

## Durham E-Theses

---

### *Identification and characterisation of behavioural genes of agrobacterium tumefaciens.*

Marashi, Sayyed Hassan

#### How to cite:

---

Marashi, Sayyed Hassan (2000) *Identification and characterisation of behavioural genes of agrobacterium tumefaciens.*, Durham theses, Durham University. Available at Durham E-Theses Online:  
<http://etheses.dur.ac.uk/4222/>

#### Use policy

---

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full Durham E-Theses policy](#) for further details.

---

Academic Support Office, Durham University, University Office, Old Elvet, Durham DH1 3HP  
e-mail: [e-theses.admin@dur.ac.uk](mailto:e-theses.admin@dur.ac.uk) Tel: +44 0191 334 6107  
<http://etheses.dur.ac.uk>

**Identification and characterisation of behavioural genes of  
*Agrobacterium tumefaciens*.**

By

Sayyed Hassan Marashi

A thesis submitted to the Department of Biological Sciences

University of Durham

In accordance with the requirement for the degree of

Doctor of Philosophy

The copyright of this thesis rests with the author. No quotation from it should be published in any form, including Electronic and the Internet, without the author's prior written consent. All information derived from this thesis must be acknowledged appropriately.



26 MAR 2002

Thesis  
2000/  
MAR

# Identification and characterisation of behavioural genes of *Agrobacterium*

## *tumefaciens*.

Sayyed Hassan Marashi

### Abstract.

Three behavioural mutants (*fla-8*, *mot-6* and *cheL*) generated by transposon Tn5 mutagenesis and localised on cosmid pDUB1905 were studied. The cosmid pDUB1905, from a representative genomic library of the *Agrobacterium tumefaciens* C58C1 chromosome has previously been partly mapped and found to contain genes concerned with flagella. In this study a region of 5860 nucleotides from a 12 kb *Bam*H1 fragment of cosmid pDUB1905 was sequenced completely in both directions. Homology searches of this sequence with sequence databases and other computer analysis revealed two flagellar-related genes (*flhA* and *fliR*), a chemotactic gene (*cheL*) and four open reading frames (*orfX*, *orfW*, *orfY* and *orfZ*) with no significant sequence identity to any open reading frame in databases. A putative promoter-like sequence was also found upstream of *orfZ*. The FlhA and FliR are the inner members of type III flagellum-specific export apparatus which are responsible for delivering the flagellar subunits lacking a signal peptide leader to the surface of the cell. These have counterparts in the type III secretion proteins system responsible for transporting pathogen proteins to host cells. The function of CheL has not yet been identified. Three ORFs have chaperone characteristics.

A mutant was created by insertion of a neomycin resistance cassette in the *fliR* homologue to determine the effects of the gene on motility. Phenotype analysis of the mutant showed no flagella and motility with small swarming pattern comparing to wild type, indicating that *fliR* is indeed a flagellar gene.

In this study two more members of flagellum-specific export, a chemotactic gene, three open reading frames which could have specific chaperon activity, and an unknown open reading frame were identified in *A. tumefaciens* C58C<sup>1</sup>.

TO:

My **mother** and **Father**,

Wife, **Hoda**,

Son **Hossien**,

**Sister** and **Brothers**

For their love and encouragement

## **Declaration**

I declare that the work contained within this thesis submitted by me for degree of Doctor of Philosophy is my own original work, except where otherwise stated, and has not been submitted previously for a degree at this or any other University.

The copyright of this thesis rests with the author. No quotation from it should be published without prior written consent and the information derived from it should be acknowledged,

## Acknowledgements.

This work would not have been possible without the support of the following people. I would like to express my gratitude to my supervisor Dr. Charlie shaw for allowing me to undertake this project and for his understanding, support and guidance he has given me during my PhD.

Too many thanks to mention to Dr. Bill Deakin and Dr. Danita Pearson for “starting the whole ball rolling” and giving me the direction and good advice.

Special thanks must go to Dr. Emma. L. Wright for her valuable advice and help especially for proof reading this thesis and encouragement whilst writing up. I consider myself extremely lucky to work with such a kind people who have made the study for my PhD, so pleasant and so much fun.

I am grateful to Julia Bartley, Vicky Kelly and Gillian storey for DNA sequencing, and all other members of Biological Sciences Department who helped me during these years.

I would like to appreciate Ministry of Culture and Higher Education of Iran and Ferdowsi University of Mashhad for providing the financial support of my study and giving me opportunity to carry out this work.

Finally I must thank my wife, Hoda who's always been there for me inspiration, love and encouragement when I have needed it most. I am also wish to thank my son, Hossein for his patience, kindness and understanding during these years. I would also like to thank all other members of my family in particular my Father, sister and brothers who have always provided encouragement throughout my years of studying, and without whom I would not have achieved so much.



## CONTENTS

<b>1. Introduction .....</b>	<b>1</b>
1.1 <i>Agrobacterium tumefaciens</i> : Characteristics and Classification .....	2
1.1.2 Soil Ecology.....	4
1.1.3 Crown Gall Tumours .....	5
1.1.4 Opines.....	6
1.1.5 Ti plasmid .....	8
1.2 Virulence in <i>Agrobacterium tumefaciens</i> .....	11
1.2.1 Host-pathogen contact .....	11
1.2.2 Plant Signals .....	12
1.2.3 Signal transduction .....	13
1.2.4 T-strand generation.....	16
1.2.5 T-complex formation.....	17
1.2.6 T-DNA transformation to the plant cell.....	19
1.2.7 T-DNA integration into the plant genome.....	22
1.2.8 T-DNA expression.....	24
1.3 Chemotaxis .....	24
1.3.1 Transmembrane receptors.....	28
1.3.2 The Histidine Kinase .....	30
1.3.3 The aspartate kinase, CheY .....	31
1.3.4 Other proteins involved in chemotaxis.....	32
1.3.4.1 CheW coupling protein.....	32
1.3.4.2 CheR methyltransferase protein.....	32
1.3.4.3 CheB methylesterase protein .....	33
1.3.4.4 The CheZ protein.....	33
1.3.5 Signal transduction in bacterial chemotaxis .....	34
a) Excitation phase.....	35
b) Adaptation phase .....	37
1.3.5.2 The Phosphotransferase (PTS) system .....	40
1.3.5.3 Aerotaxis/ Energy taxis in <i>E. coli</i> .....	40
1.3.6 Other factors regulating chemotaxis.....	41

Intracellular Ca <sup>2+</sup> levels .....	41
Intracellular fumarate level .....	41
The acetate effect .....	42
1.3.7 Chemotaxis in other bacteria .....	42
1.4 Motility in <i>E. coli</i> and <i>S. typhimurium</i> .....	43
1.4.1 Flagellar Structure .....	44
1.4.2 Chromosomal organization of the flagellar and chemotactic genes of <i>E. coli</i> and <i>S. typhimurium</i> .....	48
1.4.3 Flagellar assembly and typeIII flagellum-specific export apparatus .....	49
1.4.4 Motility and Chemotaxis in <i>Agrobacterium tumefaciens</i> .....	55
1.5. Analysis of chemotaxis and Motility genes in <i>Agrobacterium tumefaciens</i> .....	56
1.6 Aims of this project .....	57
<b>2. Materials and Methods. ....</b>	<b>59</b>
2.1 Materials. ....	60
2.2 Bacterial Strains and Plasmids. ....	61
2.2.1 <i>E. coli</i> Strains .....	61
2.2.2 <i>A. tumefaciens</i> Strains .....	61
2.2.3 Plasmids .....	61
2.2.3.2 Plasmids Used in Mutant Construction. ....	62
2.2.3.4 Recombinant SK+ Plasmids Containing Subcloned <i>Agrobacterium</i> Behavioural Genes .....	63
2.3 Bacterial Growth Media, Conditions and Procedures. ....	63
2.4 Isolation of DNA. ....	65
2.4.1 Alkaline Lysis Plasmid Minipreps .....	65
2.4.2 Larger Scale Plasmid Preparation .....	66
2.4.3 Small Scale Preparation of Bacterial Chromosomal DNA .....	66
2.4.4 Small Scale Preparation of Bacterial Chromosomal DNA - 2. ....	67
2.5 DNA Manipulations. ....	67
2.5.1 Phenol:Chloroform Extraction of DNA. ....	67
2.5.2 Restriction Endonuclease Digestions. ....	68
2.5.3 Agarose Gel Electrophoresis. ....	68
2.5.4 Isolation of DNA Fragments from Agarose Gels Using Silica Fines .....	70
2.5.5 Filling in 3'-Recessed Termini .....	70
2.5.6 Ligation of DNA .....	71

2.5.7 Polymerase Chain Reaction (PCR).....	71
2.5.7.1 Hot Start PCR.....	73
2.6 Transformation of <i>E. coli</i> .....	73
2.6.1 Preparation of Competent Cells.....	73
2.6.2 Transformation Procedure.....	74
2.7 DNA Hybridisation Procedures.....	74
2.7.1 Radio-Labeling of DNA Fragments.....	74
2.7.2 Southern Blotting.....	74
2.7.3 Hybridisation of Radio-Labelled Probes to Blots.....	75
2.7.4 Washing of Probed Blots.....	76
2.7.5 Detection of Hybridising Probes.....	76
2.7.6 Stripping of Probed Blots.....	76
2.8 DNA Sequencing.....	77
2.9 Conjugation of Plasmids into <i>Agrobacterium</i> .....	77
2.10 Mutagenesis.....	77
2.10.1 Gene Replacement Mutagenesis.....	77
2.11 Bacterial Growth Measurement.....	79
2.12 Microscopy.....	79
2.12.1 Light Microscopy.....	79
2.12.2 Electron Microscopy.....	79
<b>3. Gene Identification .....</b>	<b>81</b>
3.1 Background of this project: .....	82
3.2 Sequencing plasmid pH1.....	85
3.3 Sequencing pH2 plasmid.....	89
3.4 Constructing and sequencing of plasmid pH3.....	90
3.5 The <i>flhA</i> homologue: .....	100
3.6 The <i>fliR</i> homologue: .....	110
3.7 <i>orfX</i> .....	117
3.8 The <i>cheL</i> homologue: .....	119
3.9. <i>OrfY</i> , <i>orfW</i> and <i>orfZ</i> :.....	122
3.10 Discussion.....	128
<b>4. Mutagenesis.....</b>	<b>135</b>
4. Mutagenesis of <i>fliR</i> :.....	136
4.1 Creation of the <i>fliR</i> mutant plasmid:.....	136

4.2 Construction of an *A. tumefaciens fliR* mutant strain: ..... 138

4.3 Phenotypic analysis of *fliR*Neo mutant:..... 143

4.4 Discussion:..... 143

**5. Conclusion .....146**

5. Conclusion:..... 147

**6. References: .....153**

## FIGURES

Figure 1.1.1 Electron micrograph of a wild type <i>A. tumefaciens</i> C58C <sup>1</sup> .....	3
Figure 1.1.5.1 Genetic map of an octopine-type Ti plasmid. ....	9
1.2.6.1 Diagram showing main steps of transfer of T-DNA to plant cell.....	20
Figure 1.2.6.2 Diagram showing the basic structure of typeIV secretion system (Ti pilus) .....	21
Figure 1.3.5.1 Diagram showing main steps of the chemotactic response in <i>E. coli</i> .38	
Figure 1.4.1 Diagram showing the general structure of the <i>E. coli</i> flagellum. ....	46
Figure 1.4.3.1 Diagram showing Hypothetical structure of typeIII Flagellum-specific export apparatus.....	52
Figure 3.1.1 Tn5 mutagenesis.....	83
Figure 3.1.2 Isolation of behavioural mutants .....	84
Figure3.2.1 Cosmid pDUB1905 and major plasmids drivated from the cosmid. ....	86
Figure 3.2.2 13.2.2 Double digestion of pAB1 with <i>Bam</i> HI- <i>Eco</i> RI , and pH1 with <i>Eco</i> RI- <i>Eco</i> RV.....	88
Figure 3.4.1 Diagram outlining the main steps involved in PCR to synthesise the beginning of <i>flhA</i> . ....	91
Table 3.4.1 some characteristic of potential genes and <i>orf</i> found in this work .....	92
Figure 3.4.2 The complete sequence of the sense strand of 5.8 kb region of the cosmid pDUB1905.....	93
Figure 3.4.3 DNA Strider open reading frame map of the sequenced region of pDUB1905 showing the located and name given to each identified open reading frame. ....	97
Figure3.4.4 Diagram of Testcode result of the sequenced region of pDUB1905.....	98
Figure 3.4.5 Strategy for the partial sequencing of pDUB1905 .....	99
Figure 3.5.1 The entire coding and untranslated regions of <i>flhA</i> homologue and the predicted gene product.....	102
Figure3.5.3 Hydropathy profiles of the <i>A. tumefaciens</i> FlhA homologue and FlhA from other bacteria, LcrD and MxiA from typeIII export system. ....	109
Figure 3.6.1 The untranslated and coding regions of <i>fliR</i> homologue with its predicted translated product.....	111
Figure 3.7.1 The DNA sequence of <i>orfX</i> with its putative translated protein and DNA sequence upstream of <i>orfX</i> .....	118

Figure 3.7.2 Hydropathy profiles of the ORFX <i>A. tumefaciens</i> . ....	119
Figure3.8.1 The DNA sequence of <i>cheL</i> with its putative protein and DNA region upstream of it. ....	120
Figure 3.8.2 Pairwise alignment (pretty box) of potential CheL of <i>A. tumefaciens</i> And <i>C. crescentus</i> . ....	121
Figure 3.8.3 Hydropathy profiles of the <i>A. tumefaciens</i> CheL homologue and its counterparts in <i>C. crescentus</i> . ....	122
Figure 3.9.1 The DNA nucleotide and deduced amino acid sequence of <i>orfY</i> with untranslated region upstream of the <i>orf</i> . ....	124
Figure3.9.2 The DNA nucleotide and deduced amino acid sequence of <i>orfW</i> with untranslated region upstream of the <i>orfY</i> . ....	
Figure3.9.3 The DNA nucleotide and deduced amino acid sequence of <i>orfZ</i> with untranslated region upstream of the <i>orf</i> . ....	126
Figure3.9.4 Hydropathy profiles of the three unknown ORFs , ORFY, ORFW and ORFZ in <i>A. tumefaciens</i> . ....	127
Figure 4.1.1 Diagram outlining the main sub-cloning steps involved in the construction of the <i>fliR</i> mutant by insertion of neomycin cassette. ....	137
Figure 4.3.1 Agarose gel of the genomic DNA cut with <i>HindIII</i> and subsequent Southern blot of nine putative <i>fliR</i> Agrobacteria mutant strain hybridised with <i>fliR</i> fragment (a) and neomycin cassette (b). ....	139
Figure 4.2.2 Diagram showing the position and approximate size of the hybridising fragments of <i>fliR</i> and neomycin cassette to C58C <sup>1</sup> and <i>fliR</i> mutant strain. ....	141
Figure 4.3.2 Photograph showing the swarming behaviour of the <i>fliR</i> mutant (right) and C58C1 (left) as a control. ....	142
Figure 4.3.3 Electron micrograph of the <i>fliR</i> mutant strain. ....	142
Figure 4.4.1 showing <i>fla-8</i> , which is unflagellated and <i>mot-6</i> mutant, which is flagellated .....	144

TABLES

Table 3.4.1 Some characteristic of potential gene and *orf* found in this work .....90

Table 3.5.1 Some of proteins with sequence homology to potential *A. tumefaciens*  
    FlhA ..... 103

Table 3.5.2 Transmembrane domain of *A. tumefaciens* FlhA. .... 104

Table 3.6.1 Proteins with homology to potential *A. tumefaciens* FliR ..... 111

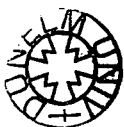
Table 3.6.2 Transmembrane domains of *A. tumefaciens* FliR. .... 112

### Abbreviations.

Amp	=	ampicillin
ATP	=	adenosine triphosphate
Bp	=	base pairs
BSA	=	bovine serum albumin
CAMP	=	cyclic adenosine monophosphate
Cm	=	chloramphenicol
DMF	=	N,N-dimethyl formamide
dNTPs	=	deoxyribonucleoside triphosphates
DTT	=	dithiothreitol
EDTA	=	ethylenediaminetetraacetic acid
Gm	=	gentamycin
IPTG	=	isopropyl- $\beta$ -D-thiogalactopyranoside
Kan	=	kanamycin
Kb	=	kilobase pairs
KDa	=	kilodaltons
MOPS	=	3-(N-morpholino)propane sulphonic acid
Neo	=	neomycin
OD	=	optical density
PEG	=	polyethylene glycol
p.s.i	=	pounds per square inch
Rif	=	rifampicin
RNAase	=	ribonuclease
SDS	=	sodium dodecyl sulphate
Strep	=	streptomycin
Tc	=	tetracycline
Tris	=	tris(hydroxymethyl)aminomethane
UV	=	ultraviolet
wrt	=	with respect to
X-Gal	=	5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside
5'	=	5' terminal phosphate of DNA molecule
3'	=	3' terminal hydroxyl of DNA molecule



# 1. Introduction



## 1.1 *Agrobacterium tumefaciens*: Characteristics and Classification

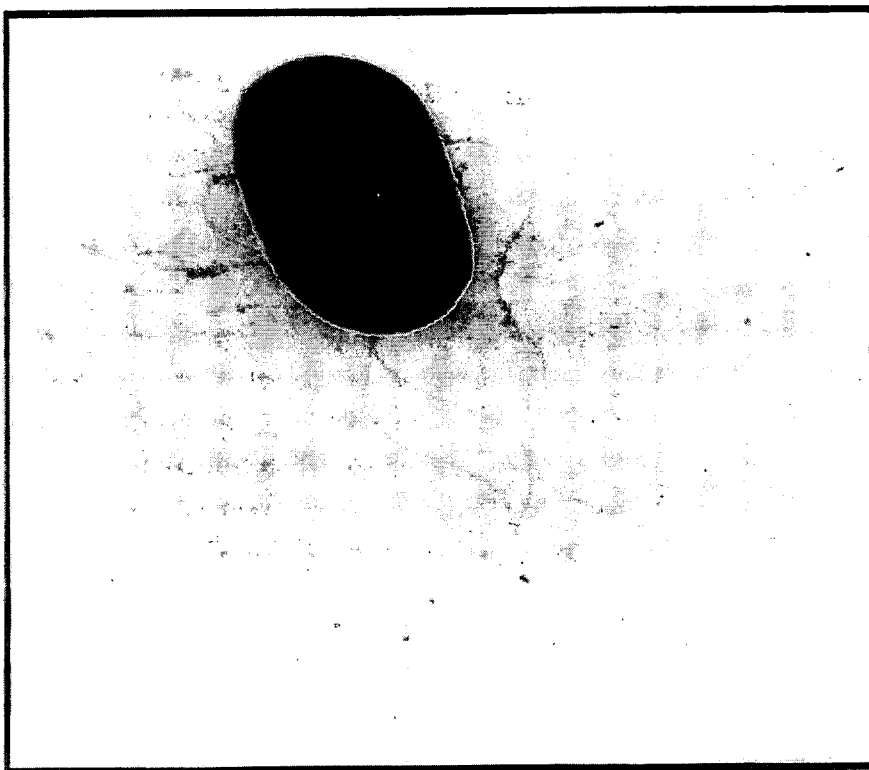
By any means *Agrobacterium tumefaciens* is one of the most successful organisms. It is one of the dominant bacterial species in the rhizosphere and it is capable of colonising and pathogenising plant hosts, causing epidemics of crown gall disease. *Agrobacteria* appear to have a great advantage in their capability for pathogenicity; making them one of the best natural vectors for introducing foreign genes into higher plants.

The most important feature of the pathobiology of this organism is its wide host range. *Agrobacterium tumefaciens* is a pathogen that can infect over 300 genera of dicotyledonous plants, at least four families of monocots and over 40 species of gymnosperms (54, 65).

*Agrobacteria* morphologically are aerobic, Gram-negative, short, encapsulated, rods, and usually 0.6-1.0  $\mu\text{m}$  wide by 1.5-3.0  $\mu\text{m}$  long. They are motile and have an unusual flagellation pattern with a polar tuft of two flagella and 2-4 lateral flagella (see fig 1.1) (97). They produce copious extracellular polysaccharide when grown on carbohydrate-containing medium. Frequently, they produce cellulose fibrils which are thinner and longer than their flagella. The colonies of *Agrobacteria* are non-pigmented and smooth. The optimum temperature for growth of *Agrobacterium tumefaciens* is 28°C.

*Agrobacterium* together with *Sinorhizobium* (formerly *Rhizobium*), *Rhizobium*, *Rhodobacter*, *Caulobacter* and *Azospirillum* are a member of the  $\alpha$ -subgroup of the class Proteobacteria. The family Rhizobiaceae contains soil bacteria which associate with plant cells and carry large plasmids.

The classifications of the genus *Agrobacterium* is divided into four species based on their phytopathogenicity (147). The pathogenic species are: *A. tumefaciens* which causes crown galls, *A. rhizogenes* which produces hairy root, *A. rubi*, the agent inducing cane galls on rubus species and *A. radiobacter* which is non-pathogenic. Each species is further divided upon the production of specific chemicals by the galls or infected roots called opines. This method of classification has been shown to be unrealistic since the phytopathogenicity of species is based on the production of opines. The genes coding for opine synthesis are on a large tumour-inducing (Ti) or



**Figure 0.1.1.1 Electron micrograph of a wild type *A. tumefaciens* C58C<sup>1</sup> cell showing flagella.** Photograph taken by Dr. C. H. Shaw.

root inducing (Ri) plasmid that can be transferred easily from pathogenic strains to non-pathogenic strains resulting in the recipient strains acquiring the pathogenic properties of the donor strains. Therefore, phytopathogenicity is not suitable for species identification. In addition, it has been shown that all known Ti plasmid contain a variety of combination of opine catabolic genes and they can catabolise more than one opine (236).

An alternative system to classify the species has recently been established using phenotypic, physiological and genotypical differences, chromosomal characteristics, comparisons of protein patterns, and other tests such as growth at different temperatures and the sensitivity to sodium chloride (99 , 100). In this system of nomenclature, the genus *Agrobacterium* is divided into three biovars (biotypes) 1, 2, and 3. The three biovars are genetically and phenotypically different from each other and they contain both pathogenic and non-pathogenic strains. The strains of biovars can clearly be separated based on differences in 16s rRNA sequences. This

method of classification is more stable and valid for species identification than phytopathogenicity.

Members of the genus all possess two chromosomes, many of which demonstrate different topologies such as having linear and circular chromosomes (96). Each biovar has also a specific genome size.

Based on the kind of opines produced in the galls, *Agrobacteria* are also classified as octopine, nopaline, and leucinopine strains.

Despite of the new system of classification, the old nomenclature is still used and will be used here. The strains used in this project are derivatives of *A. tumefaciens* C58 which possesses four replicons: one circular and one linear chromosome (which is smaller) and two plasmids. This strain contains the Ti plasmid pTiC58.

### 1.1.2 Soil Ecology

*A. tumefaciens* can be isolated from both cultivated and uncultivated soils around the world and especially found in association with the rhizosphere. The numbers of bacteria present around the roots are vastly greater than those found in nearby soil (148). Pathogenic strains are usually only isolated from galls and soils surrounding the infected roots. Generally the populations of avirulent bacteria are much higher than the virulent strains in the soil (98). Since the Ti plasmid, the means of pathogenicity, is easily lost without selection, however many reports have shown that certain factors can affect the proportion between virulent and avirulent strains. In an experiment carried out to isolate *Agrobacterium* strains from a field with loamy soil that had not been cultivated for five years, the virulent number of *Agrobacteria* were much higher than expected (31). It was proposed that the loamy soil possibly increased the number of pathogenic strains according to its capacity for water retention and acid pH. The presence of weeds could also have a positive effect on virulent strain populations. Data from experiments on other related bacteria suggests that the soil type affects survival of the bacterial strains, the stability of plasmids, motility, growth and pathogenicity (31, 91). The number of virulent strains also can be affected by opines, which can be used by *Agrobacterium* but not by the host plant or majority of other microorganism. This capability of *Agrobacterium* to utilise

opines provides an environment favouring growth of the pathogen and transfer of the Ti plasmid. This hypothesis, is called the opine concept (175).

### 1.1.3 Crown Gall Tumours

The crown gall is characterised by uncontrolled and rapid plant cell proliferation at the wound site following infection by *Agrobacterium*. In nature invasion by *Agrobacterium* usually occurs at ground level near the junction of the stem and roots. The size of galls may vary in diameter from pea-size to 5-15 cm and can be up to 22.7 kg in weight (54).

The gall is composed of a mass of undifferentiated cells resulting from unlimited division and an increase in size of the vascular and parenchyma tissue (146).

Gall formation by *Agrobacterium* on dicot plants requires wounding which produces a wound meristem. Deeper wounds facilitate infection and usually result in large galls. Development of a tumour in a particular species of plant is dependent on the type of Ti plasmid possessed by the pathogenic *Agrobacterium* strains.

The cells of the gall synthesise specific compounds formed by condensation of an amino acid, a keto acid or a sugar and generically called opines which are absent from non-transformed cells. These compounds are specific growth substrates for *Agrobacterium*. Since growth of gall tumour is a benefit to virulent bacteria, the process has been named "genetic colonization".

After the induction of a tumour by the pathogenic *Agrobacterium*, the tumour cells continue to proliferate in the absence of bacteria. Also tumours induced by *Agrobacterium*, carrying wild -type Ti plasmids generally remain undifferentiated following transfer to culture. Analysis has shown that gall cells contain auxin and cytokinin sufficient for growth and cell division (54, 239, 242). The level of indole acetic acid in tumour cells is 2-500 fold higher than in non-transformed cells. The ratio between auxin and cytokinin in tumour tissue is likely to be a major determinant of the gall's phenotype. The proliferation of transformed plant cells results mostly from the expression of *iaa* and *cyt* genes, located on the Ti plasmid and responsible for the biosynthesis of auxin and cytokinin respectively.

### 1.1.4 Opines

One of the most significant characteristics of tumour cells is that they synthesise low molecular weight metabolites called opines. More than twenty opines have been identified, comprising nine structural families. Most opines are condensation products of amino acids with hexose sugars or with carbonyl compounds. A crown gall tumour synthesises only a limited number of opine molecules.

Based on the kind of opine produced by tumours, *Agrobacterium tumefaciens* strains are classified into different groups, octopine and nopaline being the most studied opine families. Octopine ( $N^2$ -[1-D-carboxyethyl]-L-arginine) is synthesised by reductive condensation of a basic amino acid such as arginine, and pyruvate within plant tumours and degraded by octopine oxidase to arginine and pyruvate. This family consists of opines: octopine, lysopine, histopine, octopinic acid and mannitylopinines (54).

Nopaline ( $N^2$ -[1,3-D-dicarboxypropyl]-L-arginine) is formed by reductive condensation of 2-ketoglutarate with L-arginine, and similar to octopine, nopaline is degraded by nopaline oxidase to arginine and  $\alpha$ -ketoglutarate. This family comprises of nopaline, nopalinic acid, and agrocinopine A and B (102).

The genes required for formation, utilisation and degradation of opines are on the Ti plasmid. According to the kind of opine synthesised by tumours, Ti plasmids are categorised into different types. Octopine and nopaline formation is carried out by octopine-type and nopaline-type Ti plasmids respectively.

Based on the effect of opines on the bacteria, opines often have been classified into two classes; the nutritional opines, which are used as carbon and energy sources, and the conjugal opines, which are used as inducer molecules for conjugation (61). Opines as energy sources are specifically metabolised by the virulent *Agrobacterium* strains, but not by most other soil bacteria or plant itself. Although the transformed plant cells produce these compounds. This ability creates an advantage for growth in the opine-rich environment by the tumour for those bacteria able to catabolise opines over non-catabolisers.

Production and utilisation of opines are binary in nature. The genes for biosynthesis of opines are located on a part of Ti plasmid called the T-DNA that is transferred to plant cells, and integrated into plant genome, driven by plant-active

promoters, while the genes that are on the Ti plasmid and allow the bacteria to use the specific opines, are activated by *Agrobacterium* promoters.

*Occ* and *noc* are the regions that encode octopine and nopaline utilisation respectively. Each region consists of different genes. The *occ* region carries the genes required for opine transportation, catabolism and also chemotaxis to opines (234, 235). Induction of these genes are carried out specifically by their cognate opines and are under control of proteins that have positive and negative regulatory functions called OccR and NocR for octopine and nopaline strains respectively (234). However the opine is not the direct inducer of these genes, involved in transport but is converted to the direct inducer during catabolism. The *occ* region encodes the regulator OccR and the transport proteins OccQ, M, P and J which allow the bacteria to uptake specific opines and then catabolise them (216).

In the octopine-type opine only one gene is required for synthesis of the opines while several genes encode enzymes involved in opine catabolism (*ooxAB* which converts octopine to arginine and pyruvate, *orcA* converts arginine to ornithine) (103).

Some opines acting as a conjugation inducer, stimulate Ti plasmid conjugation by enhancing expression of the *tra* gene which in turn activates other *tra* genes that are necessary for conjugation converting non-pathogenic strains into pathogenic strains. Opines, therefore seem to influence bacterial communities in the immediate vicinity of the tumour.

It has been shown that opines can translocate from tumour to non-transgenic part of the plant and released into the rhizosphere (175). This exuded opine can then be utilised by those *Agrobacteria* that have not induced the tumours but are present in the rhizosphere or on a part of the plant far from the site of the gall. Interestingly, certain soil isolates of *Pseudomonas* and Gram-positive coryneforms can utilise opines as sole sources of carbon and energy as well (175).

There is some evidence showing that opines can be beneficial to the galled plant in plant protection via antagonistic effect on insects and allelopathy against weed seeds (175).

Some opines are chemoattractants, and this chemotaxis is specific to the type of Ti plasmid present in the cell (103). Functions for chemotaxis require the regions encoding catabolism of that opine. *occ* and *noc* loci confer chemotaxis to their cognate opines. Chemotaxis to opines might be important in the interaction between

Agrobacteria and their plant hosts, and can play an important role in determining the distribution of Agrobacteria in the tumour. Opines play a fundamental role in recognition of plant hosts.

### 1.1.5 Ti plasmid

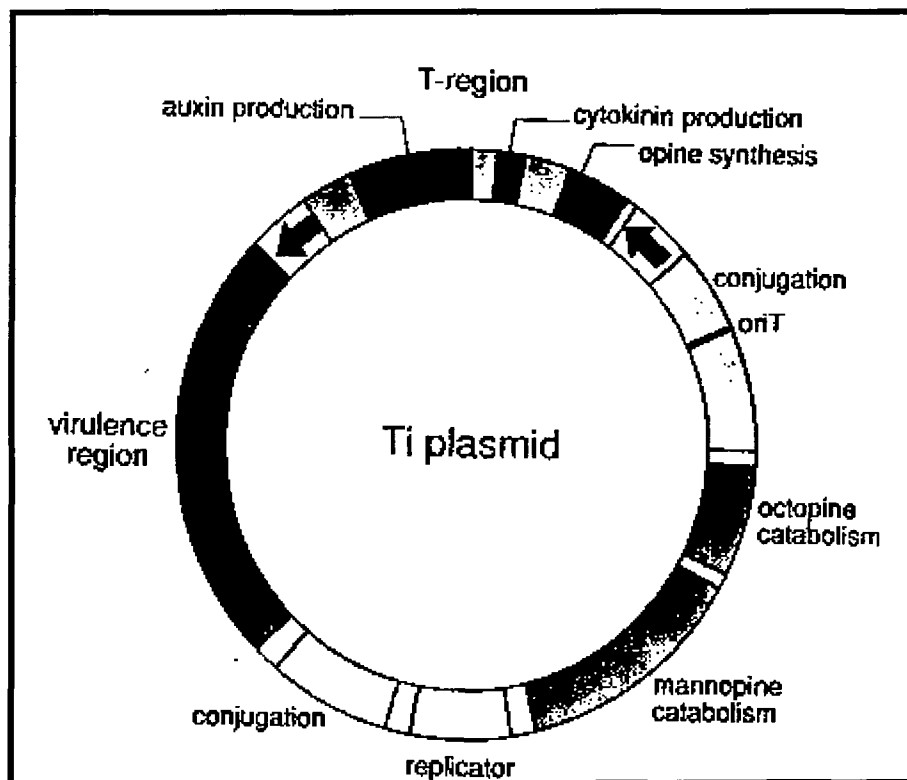
Ti plasmids are extra chromosomal elements whose host range is restricted to the rhizobiaceae. Ti plasmids are generally 190-240 kb in size and present at a low copy number (1-3 copies per cell) but can be transferred to other bacteria (90). Although tumourigenic strains usually possess Ti plasmid but many non-tumourigenic *Agrobacterium*, possessing plasmids, have also been isolated from soil (140).

Pathogenic *Agrobacterium tumefaciens* carries a specific Ti plasmid. According to the kind of opines produced by tumours, there are five Ti plasmid types: 1-Octopine type, 2-Nopaline type, 3-Agropine type, 4-Succinamopine type and, 5-grapevine type(p) (54).

Despite differences between Ti plasmids, they each consist of five distinct areas including three regions essential for tumourogenesis of bacteria (see fig. 1.4.1). The regions are *vir* region, T-DNA region and opine catabolism region plus a region, for conjugation containing *tra* and *trb* loci, and the *rep* region for replication incompatibility (see figure 1.1.5.1). The plasmid also contains genes for conjugation stimulated by opine, and sensitivity to the bacteriocin agrocin 84 in the nopaline and succinamopine plasmids. Nopaline plasmids, such as pTiC58, also have a T-DNA-independent locus, *tzs* which encodes an enzyme that produces the phytohormone cytokinin in the bacteria under *vir*-inducing conditions.

One part of the Ti plasmid is occupied by T-DNA which contains the genes that are responsible for tumour formation (oncogenes) on the infected plants. T-DNAs of different Ti plasmids have different compositions. Nopaline Ti plasmids carry T-DNA as a contiguous stretch of ~ 22 kb, whereas the T-DNA in an octopine-type Ti plasmid is composed of three independently transported segments that are left (TL) with 13 kb, central with 1.5 kb, and right (TR) with 7.8 kb size (54). The borders that flank and define the transferred region to the infected plant cells consist of a conserved 25bp imperfect direct repeat sequence (86, 231). They serve as a target for cleavage by the VirD2 endonuclease and play an important role in T-DNA transformation (233). Experiments showed that the T-DNA borders can be replaced





**Figure 1.1.5.1 Genetic map of an octopine-type Ti plasmid.** Figure adapted from (124)

by a bacterial origin of transfer (239). Mutation analysis showed that the right border is absolutely required for *Agrobacterium* pathogenicity, whereas the left border is not (179). Inversion of the right border has been shown to reduce virulence and result in transfer of nearly the entire Ti plasmid instead of the T-DNA region (239). These results suggest the polarity of T-DNA transformation from right to left, as determined by the orientation of the T-DNA border repeats. On the other hand, only the TL region is oncogenic and contains all the genes necessary for tumour formation, while the TR region has no genes for phytohormone production and is not oncogenic. Interestingly, the T-DNA itself (the sequences between the borders) has no effect on the efficiency of transfer (181, 239). Consequently, T-DNA sequences can be replaced by DNA sequences of interest for engineering plants. Now a days, non-oncogenic (disarmed) Ti plasmids are used widely as vectors for genetic transformation.

The T-DNA contains 8 to 13 genes, including a set for production of phytohormones, which are responsible for formation of the characteristic tumours on infected plants. These genes have features of plant genes rather than bacterial genes; e.g. they contain polyadenylation signals, are transcriptionally controlled by plant promoter and activator sequences and can be expressed in most dicot and some monocot plants. These genes encode enzymes which catalyse the initial steps in the conversion of adenine to zeatin and tryptophan to auxin, two plant hormones critical for cell growth, and also convert amino acids to opines. The TL region of octopine-type Ti plasmid contains the genes *auxA* and *auxB* that are responsible for auxin biosynthesis, *cyt* encodes cytokinin biosynthesis, and *ocs* controls octopine synthase. The TR region of these plasmid lacks the genes coding phytohormones and contains genes responsible for synthesis of further opines such as agropine and mannopine. In some cases, T-DNA transfer is enhanced by the so-called "overdrive sequence", a 24bp sequence, located to the right of the right border in octopine Ti plasmids.

The *vir* region which is 35-40 kb, consists of eight *vir* operons, *virA* to *virG*. Unlike octopine Ti plasmids, the nopaline type lacks the *virF* and *virH* operons but instead contains another accessory gene called *tzs*. The product of these genes, termed virulence (Vir) proteins, encode proteins involved in the detection of compounds that are released from wounded plants in the rhizosphere and mediate transfer of T-DNA into the host cells. Mutations in the *virA*, *B*, *D*, and *G* operon abolish tumour formation on all plant species, while mutation in the other *vir* genes reduces tumourgenicity and restricts the range of plant hosts.

The opine catabolism region, *occ* region, consists of genes that allow virulent *Agrobacterium* to take up opines and utilise them. The region contains the genes that are involved in the transportation and metabolism of octopine. These genes are under the control of an opine-inducible promoter, activated by the product of *occR* acting as a positive regulator. Likewise octopine type Ti plasmids, a similar set of genes are on the nopaline type Ti plasmid in the *noc* region. Another gene, *ons*, encodes a protein required for transportation of octopine across the plant cell membrane. Chemotaxis to opines is also carried out by function of the opine catabolism operon. In nopaline-type *A. tumefaciens* *noc* also confers chemotaxis to their cognate opines.

Replication and incompatibility region also maps to the Ti plasmid. In terms of compatibility, there are two Ti plasmid groups: *incRH-1* and *incRH-2*. Another

gene which is responsible for replication is *ein*. All octopine plasmids belong to the same *ein*-group, whereas nopaline plasmids belong to one of two different groups.

## 1.2 Virulence in *Agrobacterium tumefaciens*

*Agrobacteria* infect their hosts at wound sites (109), which exude specific compounds into the rhizosphere. These chemicals switch on the virulence engine of pathogenic *Agrobacterium tumefaciens* which possess a Ti-plasmid to infect the plants. Although, wounded cells are necessary for bacterial attachment, however there is evidence showing that *Agrobacteria* are able to interact with unwounding plant cells and eventually to transform them (67). *Agrobacterium* can enter leaves using their stomata. The normal invasion process occurs in three levels. Those carried out in the rhizosphere, resulting in attachment of the bacteria to plant cells, those that occur in *Agrobacterium* itself, including preparation of the T-DNA for transformation, and finally the processes that happen in the host plant, leading to transformation and integration of the T-DNA into the plant genome. The virulence process can also be divided into different steps, as discussed in the next section.

### 1.2.1 Host-pathogen contact

Cell-cell contact between pathogen and host cells is necessary for many diseases including gall tumour formation. The first step in plant infection is attachment of pathogenic *Agrobacterium* to the surface of wounded plant cells. Non-attaching mutants are avirulents (54, 115).

The attachment process consists of two stages:

1. Direct attachment of bacteria to plant cells. This initiates a loose binding and is mediated by a bacterial  $\text{Ca}^{2+}$ -binding protein called Rhicadhesin. This protein is a major determinant of attachment and virulence.  $\text{Ca}^{2+}$  plays a role in anchoring, stabilizing and for activation (130, 186).
2. Anchoring of bacteria to each other and to the plant cell surface using bacterial cellulose filaments, cellulose fibrils, stabilizes the initial binding. This generates a tight irreversible binding between the bacteria and the host cells (129). In this stage other bacteria not directly attached to the plant cell, create a large aggregation of bacteria and fibrils around plant cell surfaces. In fact only a small fraction of the

bacteria are directly bound to the host cell. The cellulose fibrils have an important role in pathogenesis. Mutants which cannot synthesise cellulose are reduced in virulence to 1/10th to 1/1000th that of the wild-type bacteria. Cellulose synthesis in *Agrobacterium tumefaciens* involves two operons of the bacterial chromosome, containing five genes (*cel* genes). Cellulose fibrils are produced at random by *Agrobacterium tumefaciens* from all sides of the cell surface (130).

Direct attachment of only a few numbers of bacteria to the host cell, suggests that plant cells could have specific receptor sites. There is evidence showing that the plant cells have a receptor of protein and pectin located on their surface, and a second binding site of unknown composition, on the plasma membrane (167). The bacteria also have a binding site composed of protein and lipopolysacchride, a component of the cell wall, located on their outer membranes (115).

The genes involved in attachment are located in the *Agrobacterium* chromosom, and include *chvA*, *chvB*, *pscA* (or *exoC*) and *Att* (181).

The success of interaction between *Agrobacterium tumefaciens* and the host cells depends on the ability of the bacteria to go through the plants defenses. In the rhizosphere, *Agrobacterium tumefaciens* also must compete effectively with other microorganisms as it attaches to the plant cell.

### 1.2.2 Plant Signals

The next step in the virulence process is induction of the *vir* genes which brings about transfer of the T-DNA into the plant cell. Activation of the *vir* genes requires the presence of specific inducers which are released from wounded plants into the rhizosphere (177). The sap from wounded plants has a characteristic acidic pH of 5.0-5.8, and contains a high amount of different aromatic and aliphatic chemicals such as phenolic compounds, which serve as plant signals, and specifically stimulate *Agrobacterium vir* gene expression. Phenolics also serve as attractants, activating chemotaxis in *Agrobacterium* (182, 187, 188). In addition, thousands of healthy somatic cells, with intact cell walls, which are released from the roots of many plant species into the rhizosphere everyday, are capable of producing extra cellular signals that induce microbial gene expression (237).

Phenolic structures alone are sufficient to induce the expression of the *vir* genes. Acetosyringone, a monocyclic aromatic hydrocarbon and  $\alpha$ -hydroxy-

acetosyringone, which are released naturally from injured plants, but are not found in undamaged plants, are examples of effective phenolic inducers. Other phenolic inducers are lignin precursors coniferyl alcohol, sinapyl alcohol, ferulic acid, and sinapinic acid (187). Some of these inducers are intermediates in cell repair biosynthetic pathways and for cell wall repair (181). So far, more than 40 low-molecular-weight phenolic inducers have been identified. Most of the monocots do not release such phenolic compounds; hence the virulence by *Agrobacterium* cannot be processed in these plants.

It is worth noting that most of the compounds released from wounded plants are bactericides at high concentrations and many also contain inhibitors of *vir* induction. Monocots also release some of these inducer compounds suggesting that *Agrobacterium* is able to recognise monocot wound sites, but further steps in the infection process might be blocked (13). Induction of the *vir* genes is dependent on a number of environmental factors including pH, presence of some phenolic compounds, presence of some monosaccharides, phosphate levels and presence of some heavy metals. The optimal conditions required by various strains for *vir*-induction are slightly different, however all require particular phenolic compounds and a pH of 5.0-5.8 (214).

### 1.2.3 Signal transduction

Detection of the plant signals that induce the *vir* genes is mediated by a two-component sensory regulatory system. The microorganisms use this kind of system to monitor their surroundings and make appropriate responses as they face high competition for limited resources and have little ability to change their environment. These systems respond, with a high sensitivity and efficiency, to changes in chemical environments around the bacteria. Changes in chemical composition are translated to signaling pathways often involving two families of signal-transducing proteins. One family is a membrane sensor protein that is responsible for recognising the changes in extracellular signals and transduced into appropriate intracellular events. These compounds surrounding the bacteria, and the other member that is a cytoplasmic protein, that transduces information from the membrane sensor to the cell. It has been suggested that a bacterium contains about 50 pairs of these sensors that regulate many of fundamental functions, such as motility and pathogenicity (197). The key function

acetosyringone, which are released naturally from injured plants, but are not found in undamaged plants, are examples of effective phenolic inducers. Other phenolic inducers are lignin precursors coniferyl alcohol, sinapyl alcohol, ferulic acid, and sinapinic acid (187). Some of these inducers are intermediates in cell repair biosynthetic pathways and for cell wall repair (181). So far, more than 40 low-molecular-weight phenolic inducers have been identified. Most of the monocots do not release such phenolic compounds; hence the virulence by *Agrobacterium* cannot be processed in these plants.

It is worth noting that most of the compounds released from wounded plants are bactericides at high concentrations and many also contain inhibitors of *vir* induction. Monocots also release some of these inducer compounds suggesting that *Agrobacterium* is able to recognise monocot wound sites, but further steps in the infection process might be blocked (13). Induction of the *vir* genes is dependent on a number of environmental factors including pH, presence of some phenolic compounds, presence of some monosaccharides, phosphate levels and presence of some heavy metals. The optimal conditions required by various strains for *vir*-induction are slightly different, however all require particular phenolic compounds and a pH of 5.0-5.8 (214).

### 1.2.3 Signal transduction

Detection of the plant signals that induce the *vir* genes is mediated by a two-component sensory regulatory system. The microorganisms use this kind of system to monitor their surroundings and make appropriate responses as they face high competition for limited resources and have little ability to change their environment. These systems respond, with a high sensitivity and efficiency, to changes in chemical environments around the bacteria. Changes in chemical composition are translated to signaling pathways often involving two families of signal-transducing proteins. One family is a membrane sensor protein that is responsible for recognising the changes in extracellular signals and transduced into appropriate intracellular events. These compounds surrounding the bacteria, and the other member that is a cytoplasmic protein, that transduces information from the membrane sensor to the cell. It has been suggested that a bacterium contains about 50 pairs of these sensors that regulate many of fundamental functions, such as motility and pathogenicity (197). The key function

of signal transduction involves phosphorylation of the membrane sensor protein that is a protein kinase.

*Agrobacteria* employ a two-component signal transduction system, composed of the virulence proteins VirA and VirG, to regulate infection of plants (223, 224, 226). Both VirA and VirG are the Ti-plasmid encoded genes required for chemotaxis and virulence induction mediated by phenolic compounds. The VirA/VirG system is thought to have multifunctional role, since at different phenolic concentrations it is known to affect both chemotaxis and Vir-induction. Low concentration (nonomolar) trigger chemotaxis alone, higher concentration (micromolar) also bring about Vir-induction and can suppress chemotaxis, thereby restricting the bacteria to the wound site and hence increasing, the infection likelihood. In *Agrobacterium* therefore, compounds released from plants trigger two specific activity, chemotaxis and virulence.

VirA is a sensor protein acting directly or indirectly as a receptor for phenolic compounds, like acetosyringone and specific monosaccharides, which are plant metabolites or monomers of plant cell wall polysaccharides (13). It has autokinase and phosphotransferase activity *in vitro*, resulting in autophosphorylation on amino acid histidine 474 (93, 226). It is an inner membrane protein, composed of four structural and functional domains: i) a receptor domain consisting of an N-terminal periplasmic domain and two transmembrane regions, designated TM1 and TM2, ii) a linker domain (cytoplasmic domain), iii) a transmitter domain that contains protein kinase activity, and iv) a phosphoryl receiver domain (40, 111, 133). Investigation into the properties of VirA have revealed that the periplasmic domain may not be the sensor domain for phenolics and direct interaction between VirA and acetosyringone requires that this metabolite diffuse through the outer membrane to reach VirA, which lacks extracellular domains (226). The periplasmic domain has a role in the detection of changing of pH and temperature of the environment. It has been suggested that VirA itself may not bind phenolics, rather plant phenolics may initially interact with chromosomally encoded protein of 10 and 21 kDa called p10 and p21. Binding to these complexes of phenol-binding proteins then activates VirA (166).

Some sugars, in particular the acidic monosaccharides D-galacturonic acid and D-glucuronic acid (both monomers of major plant cell wall polysaccharides) and other sugars such as glucose, arabinose, and galactose enhance *vir*-induction when acetosyringone are absent or present in small amounts (<10 $\mu$ m) in the environment.

The product of the chromosomal *chvE* gene is required to interact with these sugars and in turn interact with the periplasmic domain of VirA (51). This reaction results in supersensitivity of VirA binding to phenolic compounds. Transcription of *chvE* is regulated by the gene *gbpR* which lies adjacent to *chvE* on the chromosome. (64) In contrast, low amount of opines that are attractants for certain *Agrobacterium*, enhance *vir* gene expression in the presence of acetosyringone in the environment. This process is Ti-plasmid dependent and requires functions encoded by the opine catabolism operons of the Ti-plasmid.

As a general rule, conditions appropriate for *vir* gene induction are unsuitable for bacterial growth, whereas conditions that support bacterial growth are insufficient for *vir* gene induction. Therefore, it has been proposed that in low opine concentrations, *Agrobacterium* stay in the optimal state for *vir* gene induction, whereas at high levels of opine (produced in mature tumours) *Agrobacterium* are able to grow and use the opines as an energy source (181).

Interaction between plant signals and VirA leads to autophosphorylation of VirA at its histidine 474 residue in the transmitter domain. The phosphohistidine high-energy phosphate bond is then transferred to the ASP-52 or possibly the ASP-8 residue of VirG. The phosphorylated VirG is very stable, this is thought to facilitate maximal levels of *vir* gene induction (181).

The VirG protein is a cytoplasmic protein, but may also be partially associated with the *Agrobacterium* membrane. It acts as a transcriptional regulator of the other *vir* genes. To activate expression of the other *vir* genes, the C-terminus of VirG interacts with the *vir*-box, a conserved 12-bp sequence, which is located in the 5'-non-coding regions of all *vir* operons including *virA* and *virG* (94, 158). Each *vir* operon possesses up to five *vir*-boxes. The absence of a *vir*-box from the *virA* promoter is one factor restricting the number of plant species on which this strain can produce tumours (215). Specific nonconserved sequences downstream of the *vir*-boxes are required for perfect VirG function. These sequences prevent interference from other transducer proteins (165). Phosphorylated VirG is thought to have a high binding affinity for *vir*-boxes. There is also a possibility that VirG-P interacts with RNA polymerase, increasing RNA polymerase efficiency to recognise the *vir* promoter, or to open the closed DNA complex (89). The presence of *vir* inducers enhances the expression of *virA* and *virG*. Phosphate starvation also upregulates



expression of *virG*. In addition, the *virG* promoter is pH-inducible through a second promoter (17, 128, 225).

It is not known whether VirG is phosphorylated before or after binding to the *vir*-box. The actual role of phosphorylation is not clear, since unphosphorylated VirG can specifically bind *vir* promoters, however *in vivo*, VirG phosphorylation is required for *vir* gene expression. Phosphorylation of VirG may increase its DNA binding affinity (83).

The levels of expression of the *vir* genes are determined by the pool size of VirG. VirG also activates *virA* and *virG* transcription and therefore forms a positive feedback loop enhancing the infection process. Regulation of the *vir* regulon is also temperature dependent. The levels of expression decrease significantly when the temperature is raised above 22°C.

In conclusion, the compounds that are released from wounded plants in to the rhizosphere act as a plant signals and inducers. These signals are detected and transduced by *A. tumefaciens* activating *vir* gene expression. Sensing and transducing of the signals are mediated by VirA and VirG members of two-component regulatory system.

#### 1.2.4 T-strand generation

The T-DNA carries all the information necessary for tumourgenesis. Preparation of the T-DNA is carried out by the *vir* genes and involves isolation of the T-DNA from the Ti- plasmid. A linear, single-stranded DNA molecule, called the T-strand is generated (181, 193). This process occurs in the bacterial cell. The T-strand is transferred into the plant cell in a form called a T-complex.

The first step of T-strand formation is isolation of the T-DNA region from the Ti-plasmid. Two VirD proteins, VirD1 and VirD2, encoded by the *virD* operon are together responsible for nicking the 25-bp border repeats that serve as nick site for cleavage of the T-DNA (193, 239). The *virD* operon contains four open reading frames. *VirD1* encodes a topoisomerase enzyme that catalyses the conversion of supercoiled DNA to relaxed DNA (72, 89, 181). *virD2*, encodes an endonuclease that recognises the right T-DNA border. These two proteins act as a specific border repeat nickase and cut at specific sites, between the third and forth base pairs of the bottom strand of the borders (181) Excision is thought to occur in a 5' to 3' direction, starting

at the right T-DNA border and terminating at the left border. This initiates a free single-stranded T-DNA molecule or T-strand. Following cleavage VirD2, but not VirD1, remains tightly attached to the 5' end of the nicked T-strand at the right border nick and to the 5' end of the remaining bottom strand of the Ti-plasmid at the left border nick (47, 235). VirD2 possibly binds via the tyrosine 29 residue (181, 218, 239). This binding is thought to protect the T-strand from exonucleolytic degradation and probably prevents closure of the nick by ligase activity. In addition the bound VirD2 is thought to act as a pilot protein and direct the T-DNA to the plant nucleus.

When the T-strand is removed, the resulting gap in the Ti-plasmid will be repaired by generation of single-stranded copies. The replacement process removes the VirD2 molecule attached to the 5' end of the left border, allowing reformation of the circular DNA molecule of the Ti-plasmid. Recent data also suggests that VirD2 may have a role in ligation of the left border (154).

The final step in T-strand formation involves a 26 kDa protein from *virC* operon, called VirC1. This protein binds specifically to the overdrive sequence, which is a natural enhancer, located adjacent to the right border repeat. Overdrive sequence functions effectively in an inverted orientation and up to about 7 kb away from the right border repeat in either directions. This sequence is necessary for optimal T-strand formation when VirD1 and VirD2 are limited. VirC1 is thought to facilitate nicking of the T-DNA border and may play a role in T-strand export (236).

### **1.2.5 T-complex formation**

There is some evidence indicating that the T-strand is transferred from the bacterium into the plant cell in a nucleoprotein form, called the T-complex. The T-complex is composed of at least three components. The single-stranded T-DNA molecule, carrying all the information for tumourgenesis, associated with two types of Vir proteins, VirD2 and VirE2 (64).

The endonuclease activity of VirD2 lies entirely within the N-terminal of the protein, since this domain of the VirD2 protein is sufficient to cut the border repeat and T-strand formation, but is unable to import the T-DNA into the host cells. It seems that the C terminus of VirD2 must be involved in a step after T-strand generation. Genetic analysis has shown that this domain contains a bipartite nuclear

localization signal (NLS) and conserved sequence called omega (DGRGG), located adjacent to the NLS. They are shown to be important, but not essential for tumourgenesis (88). The omega sequence may interact with other Vir proteins involved in transport of the T-DNA-VirD2 complex into the plant nucleus or may involve in the integration process (32).

The *virE* operon encodes a 7.0 and a 60.5 kDa polypeptide in octopine Ti-plasmids and a 64 kDa polypeptide in nopaline Ti-plasmid. VirE2 is a single-stranded DNA binding protein that attached at its 5'-end to the VirD2 protein. It is thought to protect the DNA against degradation by bacterial and plant exonucleases and endonucleases during transformation, and shape the T-complex into a thin and unfolded structure (transferable form). It may also provide specific targeting signals. Based on electron microscopy data, the nopaline-specific T-complex is proposed to be 3600nm long and 2nm wide. This suggests that the T-complex would contain approximately 600 molecules of VirE2 (47) and only one of VirD2, with a molecular mass of 50,000 kDa (181). It was reported that ssDNA-binding protein, AcvB (or its functional homologue VirJ), may also associates with the T-complex (59)

It is thought that the T-complex is formed in *Agrobacterium* and then all components of the T-complex together transfer to a recipient plant cell. The T-complex model proposes that the 5' end of the T-strand is associated with the VirD2 molecules, whereas the 3' end probably has a VirE2 protein attached in its proximity. It has also been suggested that VirE2 and the T-strand are transported independently from *Agrobacterium* into the plant cell, and that VirE2 functions inside the host cells (49, 77, 181).

Recently, data indicated involvement of another VirE product, named VirE1, in transportation of VirE2 from the bacterium into the plant cell. VirE1 acts as an export chaperone for VirE2 (58, 202). VirE1, which is located adjacent to its targeted protein with no identified enzymatic activity, belongs to a new class of chaperones required for the export of virulence factors via a typeIII secretion pathway. It shares a number of properties with the typeIII secretion chaperones (for more details about chaperones see section 1.4.3 and Deng, 1999). This small protein is specifically required for the export of the VirE2. VirE1 prevents VirE2 to bind to T-strand by forming VirE2-VirE1 complex (24), inhibits VirE2 from aggregation, enhances the stability of VirE2 and perhaps maintain VirE2 in an export-competent state. This means that within *Agrobacterium*, VirE2 may not bind to the T-strand to form the T-

complex; rather may VirE2 enter the plant cell with assistance from VirE1 independently from the bacterium and formation of the T-complex and then bind to the T-strand-VirD2 complex. Another possibility is that VirE1 may facilitate export of the entire T-complex (77, 110, 202).

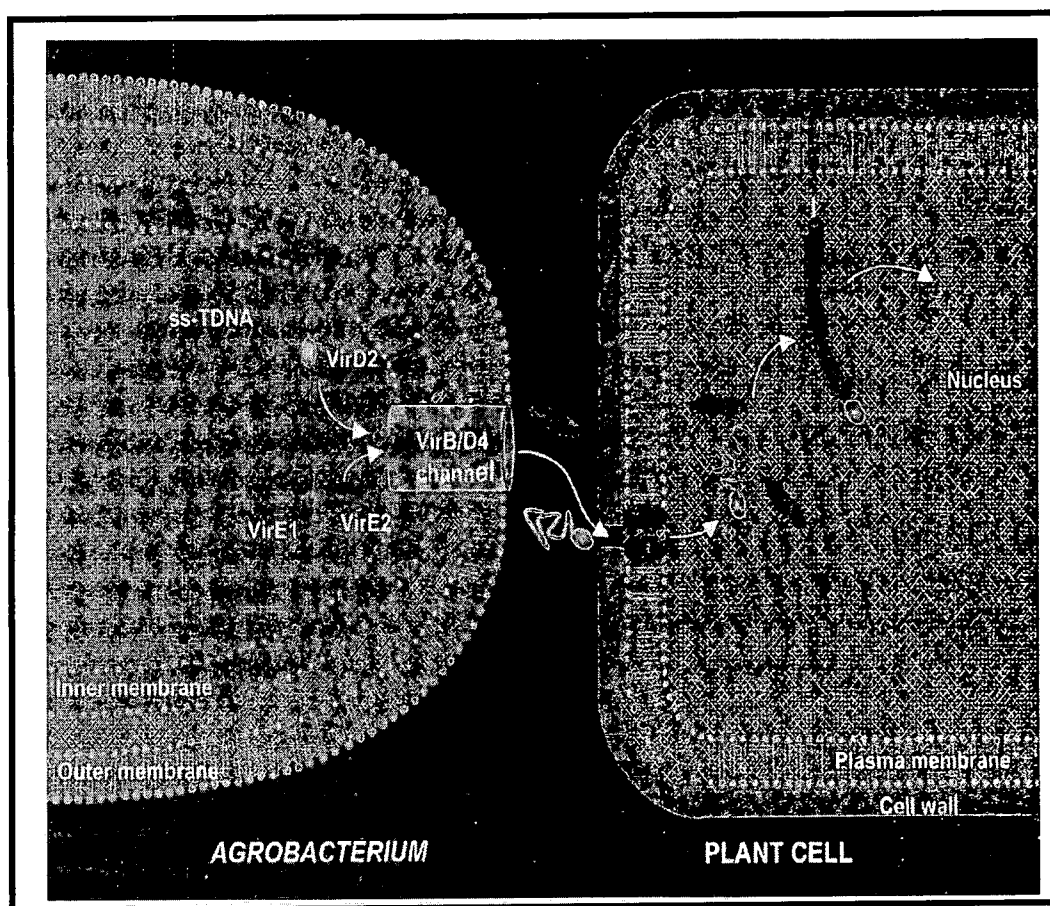
### 1.2.6 T-DNA transformation to the plant cell

It is not clear exactly how the T-DNA leaves the bacterium, crosses the membrane barrier, and enters the plant cell. Transportation of the T-DNA from donor bacterium into recipient cells (see fig 1.2.6.1) requires a passageway between the two cells. It is thought that the T-complex crosses the bacterial membrane and plant cells through a channel built by protein-protein connection of *vir* gene products (see fig. 1.2.6.2). It seems that some of the proteins encoded by the *virB* operon are involved in this bacterium-plant cell channel. The *virB* operon is the largest *vir* operon and contains 11 genes (44). It has recently been shown that the *vir* genes belong to a family of export system, called typeIV secretion system. The function of this system, which have members in  $\alpha$ ,  $\beta$  and  $\epsilon$  subgroups of proteobacteria, is direct transport of protein or DNA-protein complex across the cell envelopes of bacteria to host cells (239). Ten of them have been shown to associate with the bacterial membrane, where they may form a structure, a pilus/pore, necessary for transfer of the T-complex. The typeIV secretion system comprises two major structural components: a filamentous appendage or pilus, and a membrane associated complex that translocates substrate across both bacterial cell membranes and possibly across the host cell membranes. Based on recent studies, it has been shown that the T-DNA transfer machinery is homologous to bacterial conjugal DNA transfer system. This has been supported by the fact that most of the VirB proteins show homology with different Tra proteins of the bacterial conjugal transfer systems. In fact *vir* genes belong to a family of export system, called typeIV secretion system. The function of this system, which have members in  $\alpha$ ,  $\epsilon$  and  $\gamma$  subgroup of proteobacteria (108), is direct transport of protein or DNA-protein complex across the cell envelopes of bacteria to bacteria or from bacteria to eucaryotic cells (35, 45).

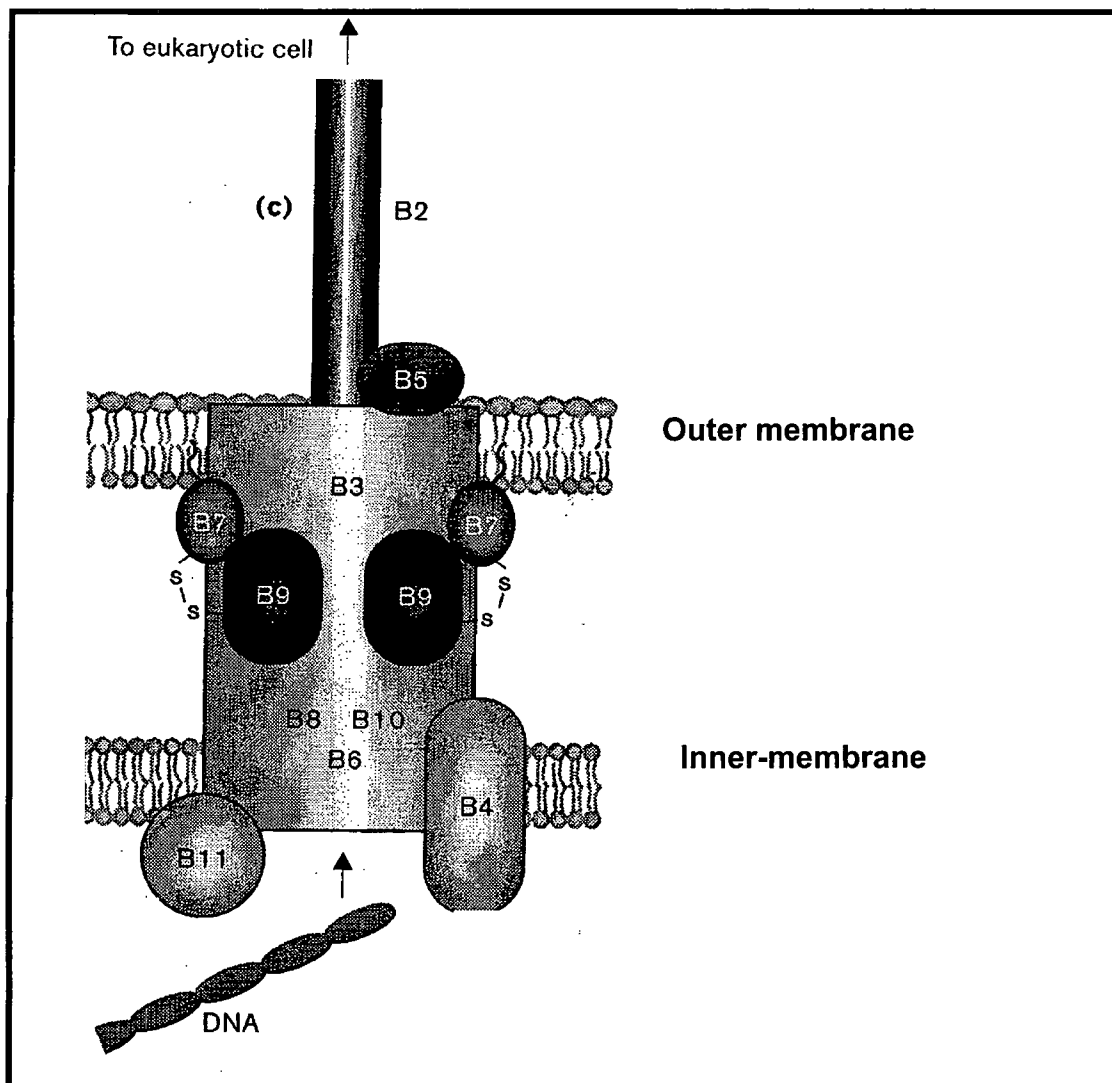
The products of two *vir* genes, *virB9* and *virB10*, are known to form a high molecular-weight complex with other membrane proteins. VirB9 is thought to stabilize or facilitate formation of a VirB10 complex. VirB7 and VirB8 stabilize

VirB9 in turn. The stabilization of some VirB proteins, in the presence of other VirB proteins, suggests that the stabilized proteins may form a multiprotein channel structure necessary for T-DNA departure.

T-complex transformation is likely to require energy. Two proteins, VirB4 and VirB11, are associated with energy supplement activity. VirB4 has a nucleotide-binding site, whereas VirB11 is an ATPase and a protein kinase, both are inner membrane proteins (44, 181). VirB1 a low molecular-weight protein, is thought to interact with the recipient plant cells facilitating anchorage of the VirB channel at the infection site on the cell surface. Recent data has shown that a protein from the *VirD* operon, VirD4, is required for T-DNA transfer and tumourgenesis. This protein which



**1.2.6.1 Diagram showing main steps of transfer of T-DNA to plant cell.** Figure adapted from (68).



**Figure 1.2.6.2 Diagram showing the basic structure of typeIV secretion system (Ti pilus).** Figure adapted from (208)

is anchored in the inner membrane of *Agrobacterium*, may play a role in the transformation of the T-complex by forming a link between the T-complex and the VirB channel. It contains a possible ATPase consensus sequence and therefore could generate the energy for T-complex transportation through the channel (114, 151).

Briefly, the T-complex is thought to be transferred from *Agrobacterium* to plant cell via a membrane pilus/pore channel (perhaps T-pilus) made up of the VirB and VirD4 proteins (108).

### 1.2.7 T-DNA integration into the plant genome

When the T-DNA enters a host plant cell T-complex form consists of the T-DNA bound with one VirD2 on its 5' end (60, 105) and hundreds of VirE2 molecules bound along its length (233, 239, 241, 242), it needs to be directed to the nucleus. It has been proposed that an endogenous plant cellular pathway is used to import the T-complex into the cell nucleus. As a rule, the active nuclear import system requires a specific nuclear localisation signal (NLS) as a targeting signal that is recognised by the plant. Since the T-strand itself does not have an NLS, the transformation of the T-strand is mediated by functions of the VirD2 and VirE2 proteins, which both contain nuclear localisation signals (48, 49, 84, 181). It has been suggested that the T-strand is directed to the nuclear pore in a polar fashion via the bound VirD2 that function as a pilot protein within the plant cell. Each of the two proteins, VirD2 and VirE2, has two nuclear targeting signals (181, 211, 239).

Some evidence has shown the accumulation of T-strands in *vir*-induced *Agrobacterium* and presence of single-stranded, but not double-stranded T-DNA in the host plant cells soon after infection (210). These results together with an experiment on properties of double-stranded and single-stranded DNA, suggest that the T-DNA enters the plant cell nucleus in a single-stranded form.

The mechanism of integration of the T-DNA into the plant genome is still not well known. Data available from different plant species indicates that the integration site is random and active transcriptional chromosomal regions are targeted. Insertion appears to be randomly distributed between translated and untranslated regions (209). T-DNA integration is not site specific, however it seems that there is a preference for integration into a transcriptionally active loci (89).

T-DNA integration is a multistep process involving several activities, such as recombination, replication and repairing. Unlike other mobile DNA elements such as transposons and retroviruses, T-DNA does not contain enzymes enabling integration. Thus T-DNA insertion into the plant genome must be mediated by proteins accompanying the T-DNA into the host plant cells and/or by host cell factors. The only bacterial proteins that are transported into the plant cells are the two polypeptides of the T-complex, VirD2 and VirE2.

Integration is an "illegitimate or random recombination" process that does not require long stretches of homology (89). Sequence homology between the incoming T-DNA and the plant chromosomal DNA has an important function in T-DNA integration. Sequence analysis showed that the left-hand end of the T-DNA (3'-end of T-strand) has a short region of homology (at least 5 contiguous nucleotides) with the insertion site of the plant chromosome, while the right-hand end of the T-DNA (5'-end of T-strand) which is attached to the VirD2 protein is restricted to just one base homology (microhomology) (209). A small deletion usually occurs at the integration site of plant DNA (less than 100 bp). Analysis of these sites has shown that the right border of the T-DNA is preserved, possibly because of VirD2 attachment to the end of the right border that protects it from the plant DNA repair machinery. However, the left border end can lose up to 1000bp. It is therefore thought that VirD2 is involved in the integration process as a ligase enzyme (239).

It has been shown that the T-DNA is inserted into the plant genome in a single-stranded form and then second-strand synthesis, is performed by the plant cell DNA repair machinery. However there is a second study of T-DNA suggesting that the T-strand is converted into a double-stranded form before integration (181).

Both models proposed that T-DNA integration starts with synapsis between plant DNA and T-DNA at the 3'-end of the T-DNA (left border), where there is a short region homology between the two DNA strands. This step is followed by the ligation of the 5'-end of the single-stranded T-DNA to the 3'-end of the plant cell DNA. The plant enzymes such as ligase perform ligation, possibly using the energy-rich phosphotyrosine bond that links VirD2 to the T-DNA. Attachment of the T-strand to one strand of the plant DNA results in a second nick on the opposite strand of the target site. Each end of the T-strand is then ligated to the plant DNA and cellular enzymes using the T-DNA as a template copy its homologous strand. A short sequence of five amino acids (DGRGG) located downstream of the VirD2 NLS, called the omega ( $\omega$ ) domain was found to be necessary for T-DNA integration (183).

Although the host cell enzymes mediate much of the integration of the T-DNA into the plant genome, a part of VirD2 and other virulence proteins may also play very important roles in this process. New evidence suggests the direct involvement of VirE2 in the recombination process. This protein is required for integration of the T-strand at the 3' end, but not at the 5' end of the T-strand molecule. In the presence of VirE2, the 3'-end is only subjected to very short deletions not exceeding 100



nucleotides. VirE2 probably physically protects the T-DNA from plant nucleases. Another possibility is that VirE2 is somehow involved in the conversion of the T-strand into a double-stranded form.

DNA rearrangements, which include duplication, deletion, inversion and base substitution have been found after integration of the T-DNA into plant DNA at the end of the insert T-DNA. This can also occur during repair and replication activities.

### 1.2.8 T-DNA expression

The integrated T-DNA genes encode enzymes that catalyse opine synthesis and its subsequent transport, and enzymes that alter the levels of plant growth regulators. Two genes, *iaaH* and *iaaM*, are responsible for enzymes that are involved in the early steps in the conversion of tryptophan to auxin, a hormone causing cell expansion. Another gene, *ipt*, provides a product that is involved in the first step of trans-zeatin and trans-ribosylzeatin production (a cell division factor) from adenine (236).

The integrated T-DNA also contains genes involved in biosynthesis of opine (104, 236). Octopine-type Ti-plasmids direct their host to produce at least eight opines (236).

The transferred genes are transcriptionally controlled by eukaryotic promoter and activator sequences and can be expressed in most, if not all, eukaryotic cells (224). The promoter activity of several T-DNA genes from *Agrobacterium tumefaciens* have been shown to be organ or tissue specific, developmentally regulated and/or influenced by plant hormones (65, 224).

The expression level of integrated T-DNA genes can be affected by different plant factors. As the T-DNA that enters the nucleus lacks a chromatin structure, besides DNA modification pattern, random methylation, mutation, rearrangements or interactions with host cell factors, occurring prior to, during or after integration, might greatly influence the expression (33).

## 1.3 Chemotaxis

As mentioned earlier (section 1.2.3), it is important for microorganisms to monitor the environment surrounding them as they face a high competition for limited resources. They use different sensitive sensory-response systems to detect and respond to the changes in the chemical or physical environments. These informations are converted to signals, which allow cells to make appropriate adaptive responses essential for cell surviving and growth. Such sensory-response systems, usually called two-component regulatory system or His-Asp kinase systems, usually consist of receptor, histidine kinase and aspartate kinase elements. Over 50 different two-component systems are operating in *E. coli* (29, 30, 38, 39, 69, 196). These systems generally control many physiological activities of bacteria including cell division, virulence, motility, adhesion, antibiotic resistance, metabolite fixation and utilization, response to environmental stress, sporulation, and different behavioural responses such as phototaxis, aerotaxis and taxis in response to cellular energy changes (69). This conserved monitoring mechanism has also been demonstrated properly in several other systems such as regulation of outer membrane porin synthesis in response to change in medium osmolarity, control of gene expression in response to the intracellular balance between carbon and nitrogen metabolism, and chemotaxis in response to attractant and repellent gradients. With such different two-component systems, specificity of kinase- response regulator interactions is crucial to avoid unwanted cross talk.

Up to present, one of the best-known two-component pathway in bacteria is the chemosensory system of *E. coli* which has become one of the leading models to describe the mechanism of the such systems. As this can provide a bright molecular picture of the chemotaxis behaviour in microorganisms, and since the principles of chemotaxis are similar in most bacteria, the chemosensory pathway of *E. coli* will be discussed in more detail.

There are at least three systems for chemoreception in *E. coli* that must link together in the central signaling pathway of chemotaxis. They are the aerotactic response and taxis to other electron acceptors (9, 10), the methylation-independent behavioural response to carbohydrates transported by the phosphoenolpyruvate (PEP) dependent phosphotransferase system (PTS) (7, 9, 205) and methylation-dependent chemotaxis (155, 199, 206).

There are two chemotaxis pathways in *E. coli*, a primitive system linked to metabolism via the citric acid cycle and a second system which is metabolism-independent with dedicated, transmembrane receptors.

Chemotaxis is the directed movement of bacterium to a component of attractant (generally nutrients such as amino acids and peptides) or away from repellent gradients (generally toxic compounds) (2). These behaviours are called positive and negative chemotaxis, respectively. The movement is carried out by flagella which rotate in either clockwise direction (CW) making random tumbles or counter-clockwise direction (CCW) producing runs (123). In the liquid media and in the absence of a chemical gradient, bacteria appear to swim in a three-dimensional random walk consisting of runs in a smooth, straight line lasting 1-2 seconds followed by tumbles during which the cell stops and randomly reorients before swimming in a new direction (20, 201). A pause seems to result from CCW to CW futile switching attempt, and there is no signal for it. In the presence of attractants or repellents, this simple swimming pattern is transferred into complex cellular behaviour. When a gradient of attractants or repellents is present, the chemosensory pathway monitors the change of the chemical concentration with time as the cell swims through different regions of the gradient. If the system detects an increase in attractant or decrease in repellent concentration with time, a signal will be sent to the motor switch to suppress CW rotation of flagella and subsequent tumbles, thereby lengthening the run towards the highest attractant concentration. Conversely, when the pathway detects a decrease in the attractant concentration or an increase in the repellent concentration with time, tumbling event will be generated by CW rotation of flagella, so that runs in this direction becomes shorter. These processes change the random walk into a walk to or away from ligands. Although such simple chemotaxis pattern is studied under controlled condition and solution media, however in natural environments the taxis on the surface of wetted substrates, likely shares some of the same switching behaviour observed for swimming in solution (69).

*E. coli* has memory, which allows detection of changes in gradient concentration. As the cell swims, *E. coli* compares the current chemical concentration that monitored by the receptor, to that chemical of the recent past as remembered by the methylation level (26).

Some compounds recognized as attractants for *E. coli* are certain amino acids, sugars, carboxylic acids, oligopeptide (1), serine and aspartate that providing the

strongest response, whereas repellents include organic molecules such as phenol, benzoate, indole, ethanol, ethylene glycol and some inorganic ions such as  $\text{Co}^{2+}$ ,  $\text{Ni}^{2+}$  and  $\text{S}^{2-}$  (125). Oxygen, fumarate and nitrates, which can act as electron acceptors in the respiratory chain, are chemoattractants (205). Chemotaxis also occurs to variation in pH and temperature (125).

The chemosensory pathway between bacteria shares nearly the same set of components, however the mechanism that operates through the system shows a great deal of variation, both between and within different bacteria. Most pathway components can be swapped between species to generate functional chimeric receptors with the expected changes to ligand specificity and without reduction in function (69).

Chemosensory proteins that communicate with one another to relay the information coming from the environment in the form of chemical signals, and regulate cellular behaviour carry out activation of chemotaxis. Three classes of proteins are involved in chemotaxis: transmembrane receptors, cytoplasmic signaling components, and adaptive methylation enzymes. Communication between the receptors and the flagellar switch motor involves four proteins which are CheA, the histidine protein kinase, CheY the response regulator, CheW, the receptor-coupling protein; and CheZ, an enhancer of CheY-P dephosphorylation. This part of the chemotaxis system works by phosphorylation reactions during intracellular signaling.

The chemosensory pathway in *E. coli* begins when the receptors sense and bind to the small-molecule ligands which enter the periplasmic compartment. The receptors, then bind to and activate the cytoplasmic histidine kinase, CheA. The receptor is coupled to the CheA by CheW protein providing a stable receptor-kinase complex. The kinase is autophosphorylated at a specific histidine residue, then the phosphate is transferred to an aspartate residue of a response regulator CheY which activates the motor switch by binding to FliM, a motor switch component. Finally, the CheZ protein provides a specific phosphatase activity which facilitates the dephosphorylation of CheY. Two other proteins, CheB and CheR which are involved in adaptation phase of chemotaxis, controls the covalent modification of the receptors by an adaptive methylation pathway (38, 69, 141, 201)].

### 1.3.1 Transmembrane receptors

The chemosensory signaling pathway in *E. coli* which works through phosphorylation, begins with sensing specific compounds in the environment surrounding the cell. Most of these signals are carried out by a set of proteins (receptors), often called MCPs (for methyl-accepting chemotaxis proteins). Some of the chemicals directly bind to the receptors but some must bind to one of the binding proteins and then, to the receptors. There is a set of at least five water-soluble binding proteins located in the periplasmic compartment of the cell that recognize some of the chemoattractants and also act as ABC transport receptors. They are: ribose-binding protein (RBP), glucose/galactose binding-protein (GBP), maltose/maltodextrins (MBP), dipeptides-binding protein (DBP), and Nickel-binding protein (in repellent response, NiBP) (69). When a chemical is bound to a binding protein, the complex of the binding protein and chemical is linked to the surface of its cognate transmembrane receptors, thereby initiating a signal process or passes the ligand to a transporter.

So far, five transmembrane receptors have been found in *E. coli* and the related bacteria, *Salmonella typhimurium* which recognize different attractants. The five chemoreceptors and their respective chemoattractants are: (a) Tsr, serine; (b) Tar, aspartate and maltose; (c) Trg, ribose, galactose; and (d) Tap, dipeptides. *Salmonella* lacks Tap but possesses Tcp, a citrate sensor and Tip (201). Trg and Tap recognize only binding proteins, Tsr only small molecules, while *E. coli* Tar is uniquely capable of sensing both small molecules and binding protein ligands. Tar can respond to MBP and aspartate as attractant and the  $\text{Ni}^{2+}$  and  $\text{CO}^{2+}$  as repellents. All of the five MCPs in *E. coli* show that the proteins are closely related and have the same structural domains. Recently, a sixth receptor involved in aerotaxis has been described, which also appears to regulate the CheA/CheY two-component signaling pathway (22). This receptor possesses a different domain organization at its N-terminus in the cytoplasm that is a suitable location for sensing the redox potential of the electron transport chain or other components of the cell interior.

Maltose, ribose, galactose, and dipeptides bind via binding proteins to their corresponding transmembrane receptors, whereas serine, aspartate, and citrate bind directly to the receptors without the assistance of a binding protein. Response to temperature, pH, and various repellents are also mediated by the receptors. It is also note worthy that each of the transducers is also responsible for movement away from

one or more repellents, however no receptor for repellents has so far been directly identified.

Each of the four soluble binding proteins is a monomer and they range in size between 30 to 51 kDa. A fifth binding proteins has also been found to detect  $\text{Ni}^{2+}$  as a repellent through Tar. All binding proteins contain a conserved two-domain structure linked by a hinge of two or three strands. This hinge is a critical feature, allowing the proteins a dynamic equilibrium between open and closed forms that can be shifted by binding the appropriate small molecules. The ligand binds in the deep cleft between the two domains. The state of cleft, opened or closed, plays an important role in binding of ligand. The closed form of the protein is the active form in both chemotaxis and transport, whereas open forms predominate when the ligand is not bound and provide a way to reduce inappropriate interaction with the membrane components. Closed forms do occasionally occur in the absence of ligand. The open form of the cleft is essential for ligand binding and release, whilst the closed form of cleft by ligand induction seems to be important for successful binding to a transmembrane receptor. Cleft closure regulates a large receptor-docking surface.

The classical receptors which are also called "MCPs" for methyl-accepting chemotaxis protein, consist of three main parts. they are : a) sensory domain, b) transmembrane domain and, c) a large cytoplasmic signaling and adaptation domain. The latter domain binds to the cytoplasmic proteins CheW, CheA and CheR generating the signaling complex. CheY and CheB also dock competitively to the complex creating a large signaling complex containing most of the pathway components except CheZ.

Each cytoplasmic domain contains four or five methylatable glutamate residues. They are activated by methylesterification and the receptors are therefore called methyl-accepting chemotaxis proteins (MCPs). There is a C-terminal called "tail" can be easily removed.

Some of the MCPs of *E. coli* and *Salmonella typhimurium* are thought to be localized primarily at one end of the rod-shaped cytoplasmic membrane called a "nose", although the cell orients this nose either toward or against the direction of swimming (126, 156).

The sensory domain of the transmembrane receptors that recognize certain chemicals is located in the periplasm at the outside surface of the cytoplasm. The periplasmic domain of the aspartate receptors consists of two 19 kDa subunits and

each of them is an antiparallel four-helix bundle. Two symmetric aspartate-binding sites are located at the interface of the two subunits (63, 69, 135). The other transmembrane chemoreceptors have the same structure, although each periplasmic sensory domain is bound to a distinct ligand.

The membrane part of the receptors, called the transmembrane signaling domain, consists of four long helices, providing by two subunits. Several experiments have shown that the transmembrane signal is carried by the second transmembrane helix (TM2) across the bilayer (41). Recent genetic studies on the aspartate receptor have revealed that the signal is transmitted largely or completely within the subunit. In fact the cytoplasmic domain of the other subunit can be removed without blocking the signal (63, 69).

The cytoplasmic domains or methyl-accepting domains of the receptors are a dimer with 31 kDa periplasmic subunit and contain the covalent modification sites for receptor adaptation and a binding site that regulates the histidine kinase by binding to CheA. It has been proposed that the dimeric cytoplasmic domain consists of three distinct functional regions: a) the linker, b) the methylation region and, c) the kinase-signaling domain (53, 69, 198). The linker region couples the transmembrane signaling helix (TM2) to the cytoplasmic domain and it is important for signal transmission. The methylation region is composed of two methylation segments. Each monomer contains four to five glutamyl residues, which are specifically methylated by CheR and demethylated by a protein glutamate methylesterase, CheB. The first segment contains the regulatory methylation sites targeted by the adaptation pathway. This segment is important for kinase regulation as it transfers the information from the periplasmic ligand-binding site to the kinase. The kinase-signaling domain forms a ternary complex with CheW and CheA and is responsible for regulation of the bound histidine kinase activity (63, 201).

### **1.3.2 The Histidine Kinase**

CheA is a key member of the two-component chemotaxis system, integrating the signals of five transmembrane sensor proteins and controlling two response regulators. CheA is autophosphorylated at His48 (23) by signaling output of the receptor and then phosphorylates CheY (132).

The kinase is a dimer of identical 71 kDa subunits with 654 amino acids, containing two symmetric active sites. Each site utilizes  $Mg^{2+}$ -ATP to drive phosphorylation of His48 (195, 201). Each subunit can be divided into four functional regions (63, 157), the first two of which fold as independent structures:

- (a) the phosphotransfer P1 domain, which possesses the reactive His48 residue that serves as the site of autophosphorylation (85),
- (b) the P2 response-regulator docking domain, which provides the binding site for CheY and CheB.;
- (c) the catalytic domain which is thought to contain all of the motifs needed to bind  $Mg^{2+}$ -ATP, recognize the P1 domain, and catalyze its phosphorylation;
- (d) the C-terminal receptor docking region which is essential for receptor-mediated regulation. Genetic studies suggest that distinct CheW- and receptor-binding sites exist within this region of CheA. The conserved C-terminal of CheA is not required for protein-protein interaction between the histidine kinase and response regulator components, nor is this region required to catalyze the transfer of phosphoryl groups from histidine to aspartate chain. Rather, the conserved kinase domain probably defines the active site for ATP binding and histidine phosphorylation.

The activity of the kinase can be increased when it interacts with the receptor and CheW.

There are two forms of CheA; the full-length molecule having 97 additional amino acids in the N-terminal domain (P1 domain) and a short form called CheAs that lacks the P1 domain and cannot be phosphorylated (157, 201).

### **1.3.3 The aspartate kinase, CheY**

In the two-component signaling pathway, the histidine kinase acts as a signal transmitter and the aspartate kinase serves as a receiver (157, 201). The receivers, including CheY, have an approximately 110-amino acid domain, referred to as a "receiver module". This domain contains the Asp-phosphorylation site.

The CheY is a 14 kDa monomer protein with 128 residues. It has two functions: as a receiver for signals from CheA by phosphorylation, and as a response regulator for motor switching. The phosphorylation site in CheY is Asp57 (170). CheY shows a great affinity for CheA when it is unphosphorylated, whereas



phosphorylated CheY exhibits a decreased affinity for CheA but a significantly increased affinity for the motor component FliM and the phosphatase CheZ (221). Interaction between CheY and FliM of the motor switch leads to a change of flagellar rotation from CCW to CW. Activation of CheY by phosphorylation generates a global conformational change that regulates the multiple docking site of CheY (15, 16, 66). These multiple docking sites interact with each of three effector proteins; P2 domain of CheA, FliM (222) and CheZ (28). The binding of  $Mg^{2+}$  carries out significant conformational changes in CheY leading to modification of the surface domain which interacts with the switch.  $Mg^{2+}$  is essential for CheY phosphorylation, dephosphorylation (CheZ dependent) (28, 195), and CheZ binding to CheY, however the CheY-P-FliM is  $Mg^{2+}$  independent. Phosphorylation of CheY leads to its dissociation from the quaternary complex, receptor-CheA-CheW-CheY (201).

### **1.3.4 Other proteins involved in chemotaxis**

There are other proteins involved in the chemotaxis response besides the two-component regulatory proteins. They act at different levels of the system and include CheW, CheR, CheB and CheZ.

#### **1.3.4.1 CheW coupling protein**

CheW is a 18 kDa coupling protein which has no regulatory or catalytic function, binds the CheA to the receptor enhancing the activation of the histidine kinase, CheA (50, 76). CheW serves, as a simple scaffolding protein and it is important for CheA function. However the specific site or residues, which are involved in receptor and kinase docking have not yet been identified. Kinase activity requires the coupling CheW, in contrast, kinase inhibition is CheW independent.

#### **1.3.4.2 CheR methyltransferase protein**

The CheR methyltransferase a 30 kDa soluble enzyme, which exist free in the cytoplasm, modifies the receptor cytoplasmic domain. It is a key factor in the adaptation phase of chemotaxis. In the cell, CheR is always found associated with the

cytoplasmic portions of the receptor complex. CheR binds to aspartate, serine (229), and citrate receptors and generates the receptor-kinase signaling complex. The receptor site which binds to CheR consists of four to five-residue sequences and located at the extreme C-terminal end of the receptor cytoplasmic domain, however the ribose/ galactose and dipeptide receptors lack this docking sequence and are hardly methylated in the absence of the aspartate or serine receptors (78).

CheR, which utilizes S-adenosyl-methionine as the methyl donor, consists of two domains. The active site cleft located between the two domains.

#### **1.3.4.3 CheB methylesterase protein**

The methyl esterase role in chemotaxis adaptation is played by CheB. This protein has two different activities (amidase and response-regulatory function), which are carried out by its different termini. The C-terminal methylesterase domain of CheB functions as an esterase to hydrolyse glutamylmethylester groups in the membrane receptor proteins (194). In some cases, CheB can hydrolyse the receptor methylation sites which are expressed as glutamines, and produces a bare glutamate side chain that is then available for methylation (69).

A linker to the C-terminal domain joins the N-terminal domain of CheB. The N-terminal domain is homologous to CheY, containing the site of phosphorylation that can be phosphorylated by transferring the phospho-His48 of CheA to its own active site similar to those of CheY (63). It seems that N-terminal domain regulates C-terminal methylesterase activity by blocking its active site physically. However, when CheB phosphorylated by CheA, generating a conformational change, this inhibition is removed, allowing CheB-P to interact with MCPs. Phospho-CheB dephosphorylates much faster than phospho-CheY by auto-hydrolysis reaction. This is because, unlike CheY-P, the phospho-CheB molecule is not dephosphorylated by CheZ (85, 194). Phosphorylation of the CheB receiver domain increases the methylesterase activity of the CheB.

#### **1.3.4.4 The CheZ protein**

CheZ protein is a component of the chemosensory pathway that is less understood. Homologues of CheZ have not been found in species outside the  $\alpha$ -

subgroup of proteobacteria (8). It serves as a phosphatase to speed up dephosphorylation of CheY and thus inactivating the tumble signals (27). CheZ can dephosphorylate CheY-P with a high specificity, either directly using a CheZ site residue or indirectly via a conformational change which stimulates autophosphatase activity of the CheY active site (69, 85). The phosphatase activity of CheZ is essential for rapid response to attractant and suppressing tumbling of bacterial cell. However, little is known about how CheZ, itself is activated.

CheZ has been isolated as a hydrophilic dimer of 24 kDa subunits. Recently a distinct functional domain within CheZ has been found. This domain contains 19 residues of the C-terminus which form part of the CheY-docking site (27). There is another possibility that the N-terminal region of CheZ may also contribute to the CheY-docking surface (173).

### 1.3.5 Signal transduction in bacterial chemotaxis

Most chemical stimuli are detected by specific receptors. Some chemicals firstly interact with a binding protein, and the complex subsequently interacts with a transmembrane receptor, however other stimuli are directly detected by receptors themselves.

There are two kind of chemotaxis in *E. coli*. In metabolism-independent chemotaxis, the signal is generated from a change in the concentration of molecules that binds to the receptors rather than their absolute levels.(112). In metabolism-dependent behaviour, the binding molecules must be metabolised to produce a signal. The term "chemotaxis " usually refers to metabolism-independent behaviour, whereas metabolism-dependent system is used for "energy taxis" which is responsible for monitoring the change in cellular energy resulting from metabolism of the signaling molecules. The metabolism-dependent system include basic requirment for bacterial growth, such as oxygen and other acceoptors, light, most sugars and a range of electron donors (7, 8).

The chemotaxis can be divided into two parts: a) the excitation phase and, b) the adaptation phase. The excitation process includes all activities which transduce the initial signal from the receptor to the motor reversing the direction of flagellar rotation. This process operates through a phosphorylation cascade. The adaptation process which follows the excitation phase, restores the flagellar rotation to

prestimulus levels, even in the continued presence of the stimulus. This process operates via a methylation mechanism.

#### **a) Excitation phase**

The nature of the signals generated by receptors and subsequently transduced was discovered approximately 10 years ago. It was shown that the signals were not electrical, but rather chemical. Chemotactic responses begin when membrane receptors, which are clustered at one of the poles, bind directly, (or indirectly with periplasmic binding proteins) to specific compounds. Binding sites for attractants or repellents and binding proteins are distinct. A single receptor sends an additive signal if more than one compound is bound. Thus maltose can block tumbling even if the aspartate binding sites on the Tar receptor are occupied. Similarly the repellent response to phenol mediated by another receptor, is abolished with addition of ribose (69).

Binding proteins contain two domains. Ligand binds in the cleft between the two domains. The cleft is open in an unoccupied state but closed when bound to a compound. Binding of a ligand to the cleft generates a contiguous docking surface and conformational changes that activate the receptor binding sites.

The sensory domains of the receptors are located in the periplasm. They recognize binding proteins or small molecule ligands. It is thought that the chemotaxis receptors transmit the signal as an intra-dimer conformational change. The second transmembrane helix (TM2), of the transmembrane domain of the receptor transduces the signal across the bilayer via the linker of the cytoplasmic region domain (41). This domain binds and regulates CheA. Recent genetic studies showed that transmembrane signals are generated completely within one subunit.

The homodimeric C-terminal domains of the receptors contain the covalent modification sites responsible for receptor adaptation, and the docking site that regulates the CheA histidine kinase. The latter site of the cytoplasmic domain of the receptor forms a ternary complex with CheW and CheA. This region is an independent functional unit containing the minimal receptor sequences required for ternary complex and activation of the kinase.

CheA is an autophosphorylation histidine kinase. It exists as a dimer, containing two symmetric active sites, which utilize  $Mg^{2+}$ -ATP to drive phosphorylation of His48. The monomeric form of CheA is not active in the

laboratory; dimerization of CheA is therefore thought to enhance kinase activity, in which one monomer phosphorylates the other (203). Phosphorylation of CheA is increased when coupling protein CheW joins the CheA-receptor complex forming a large complex comprising six subunits. Genetic studies suggest that CheW directly binds receptor during the excitatory phase (116), however CheW can form a complex with CheA which increases the rate of phosphorylation and affinity for CheY (76, 134).

The kinase uses ATP to phosphorylate itself at a histidine residue. The autokinase activity is reduced when attractant is bound to the ligand-binding receptor and increased when no attractant is present. Repellents stimulate the histidine kinase activity and speed the production of CheY-P, whereas attractants inhibit the active ternary complex and slow phospho-CheY formation, thereby raising or lowering the steady state tumble signal, respectively.

CheA provides the docking sites for CheY and CheB. These binding sites, which are located in the P2 domain of CheA, allow transfer of the phosphoryl group from the histidine kinase to the aspartate kinase protein CheY or CheB. Transmembrane signaling occurs solely within the complex is composed of a receptor dimer, CheA dimer and two molecules of CheW.

CheY acts as a phosphorylation-activated switch that controls the direction of flagellar rotation. As a result it is known as the tumble signal (176). Its action is modulated by two opposing reactions: creation of CheY-P by the receptor-kinase complex, and destruction of CheY-P by hydrolysis of its phosphate. CheY binds the receptor-CheA-CheW complex by docking to the P2 domain of CheA. Transfer of a phosphoryl group then occurs, from CheA to Asp57 of CheY.  $Mg^{2+}$  is required for phosphoryl transfer to CheY (113, 195). Phosphorylation of CheY results in a significant conformational change (18) that affects its interaction with other proteins, dissociating it from CheA, while facilitating binding with CheZ. CheZ is an auxiliary protein which accelerates the dephosphorylation of CheY (27). CheY-P binds to and regulates the flagellar motor switch (15, 16), altering flagellar rotation from the default direction of CCW to CW rotation, thereby favoring tumbling. CheY-P is therefore called "tumble factor". Mutation studies suggest that tumbling is modulated by a specific conformational change at the motor-docking surface.

After dissociation of CheY-P from the complex, the rest of the complex, containing phospho-CheA, can also phosphorylate the CheB methylesterase, thus controlling the adaptation phase of the system.

Dephosphorylation of CheY and termination of the CheY-switch interaction is enhanced by CheZ. Interaction between CheZ and CheY-P is higher than CheY itself and the binding of CheZ to both CheY and  $Mg^{2+}$  enhances CheY-P as well. Phosphorylation and dephosphorylation of CheY requires the presence of the cation  $Mg^{2+}$ . It has been proposed that CheZ, rather than catalyzing the hydrolysis of the CheY-P, acts allosterically to stimulate the auto-dephosphorylation activity of CheY-P. Some studies suggest that CheZ is not always necessary for chemotaxis to occur, and the requirement for CheZ depends on the attractant gradient background of the bacteria. Although phosphorylation increases binding of CheY to the switch, some observations have indicated that CheY phosphorylation alone is neither sufficient nor essential for binding to the motor switch. All of the current evidence is consistent with the view that CheY phosphorylation induces a conformational change that causes CheY to bind to flagellar motor proteins and thereby induce a tumble response. It has been proposed that for the generation of CW rotation an additional cytoplasmic component, which is not a chemotaxis protein, is required. Bourret *et al.* suggest that switching from CCW to CW involves at least two steps which CheY phosphorylation is the first step (184). The second step is possibly carried out by lys-109 of CheY in an event following phosphorylation.

The chemosensory pathway is extraordinarily sensitive. As little as a single attractant molecule can trigger a detectable motor response. Each *E. coli* cell contains approximately 25000 receptor monomers, 2500 molecules of CheA and equivalent amount of CheW, and 12000 molecules of CheY (200). Thus cells can exhibit a significant response when attractant ligands bind to only a few receptors. The motor is extremely sensitive to very small fluctuations in CheY-P.

## **b) Adaptation phase**

In most signal transduction pathways each reaction has a counter reaction. These activities that lead to prestimulus behaviour, have termed adaptation, habituation, etc.

The adaptation part of the chemotaxis system adapts the cell to a stimulus concentration so that the system can react even in a large amount of attractant or



Methylation of the receptors counterbalances the effects of ligand binding that diminishes the response in the presence of a constant stimulus.

CheR binds tightly to the C-terminus of some receptors (aspartate, serine, and citrate receptors), and the glutamate residues of MCPs are methylesterified via S-adenosylmethionine as the methyl donor (43). The resulting methylesters, which increase the kinase signal activation of the receptor, are hydrolyzed by the action of CheB (164). CheB is a target for phosphorylation from phospho-His48 of CheA and CheB-P removes methyl groups from the methylated receptors. In steady state, methyl addition by CheR balances methyl removal by CheB-P causing run-tumble behaviour. The feedback control of the adaptation system is generated by the receptor-kinase complex, which phosphorylates and activates CheB. CheR appears to work at a constant rate, whereas the activity of CheB is regulated by phosphorylation (7). When an attractant binds to a receptor and inhibits CheA activity, the level of CheB-P drops due to autohydrolytic dephosphorylation, resulting in a gradual increase of CheR activity. Eventually the rising level of methylation will stimulate CheA activity, thereby restoring the ability of the receptor to stimulate CheA and bring the signal to its basal level. Thus the adaptation phase of chemotaxis, exactly balances the signals generated by multiple attractants and repellents.

Attractants increase the methylation levels of the MCPs. A highly methylated transducer signals the motor to bring about CW rotation. Thus following the excitatory signals, increased levels of methylation return the receptor to a null signaling state. Conversely, repellent adaptation causes loss of methyl groups, with unmethylated receptors sending signals for CCW rotation. The level of MCP methylation also provides a simple chemical memory to find out whether the current direction of swimming is favourable or not. This process works based on a time difference between fast MCP conformational changes and relatively slow methylation changes after binding of ligands. The methylation level depends on the attractant or repellent concentration. For example the methylation level is high if the concentration was high. When the cell swims, it compares the current chemical environment, as monitored by the ligand occupancy of the receptor population, to the chemical environment of the recent past, as remembered by the methylation level. Thus if a bacterium with transducers in the adapted state moves in a long run down an attractant gradient, the still methylated, but no longer ligand bound receptor biases the flagella rotation and produce tumbling. Only if no more attractant molecules bind, to



the proteins become demethylated and return to the unstimulated state. For more reviews see (3, 9, 63, 69)

### 1.3.5.2 The Phosphotransferase (PTS) system

Prokaryotes such as *E. coli* possess a variety of systems besides the basic chemotaxis pathway, which contribute to the control of swimming behaviour. One of such systems is the phosphoenolpyruvate (PEP)-dependent phosphotransferase system, responsible for detection and translocation of a large number of carbohydrates, such as glucose and fructose, that can also act as chemoattractants (160). This system is thought to be methylation-independent and not carried out by the MCPs. There are several enzymes, EI, EII, and EIII involved in translocation and phosphorylation of carbohydrates. Phosphorylation of compounds depends on enzyme phosphorylation by a pathway involving transfer of a phosphoryl group from PEP to the enzymes. PTS phosphate moves from PEP to enzyme EI, to HPr (histidine protein) to an EII/ EIII complex and finally to the sugar.

The system requires the two-component proteins CheA, CheW, and CheY (168). It has been proposed that this system is integrated with the MCP pathway at the point of CheA (120), by phosphorylation of EI enzyme. Phospho-EI inhibits CheA autophosphorylation and hence also CheY, leading to the chemoattractant swimming response.

In spite of the importance of this system for bacteria, the chemotaxis response to PTS attractants is much weaker than chemotaxis that is carried out by the MCPs (220).

### 1.3.5.3 Aerotaxis/ Energy taxis in *E. coli*

An additional receptor has been identified in *E. coli* that responds to oxygen and a variety of other stimuli which cause changes in the energy (redox) levels in the bacteria. This receptor, referred to as Aer (since it is responsible for the aerotactic response) (22), can also activate flagellar rotation. It has been shown that Aer can sense redox potential and internal energy in the cell, in addition to external oxygen (78, 207).

The Aer receptor is located at the cytoplasmic surface of the bacterial membrane. It acts as an intracellular sensor with two cytoplasmic domains but no periplasmic sensing domain. The second cytoplasmic domain is thought to be the sensing domain. The carboxyl-terminal domain of Aer is very similar to that of a classic MCP. In contrast the amino-terminal domain looks like the redox-sensing domain.

Aerotaxis has been shown to require CheA, CheW, and CheY to function (168). Aerotaxis can also integrate with chemotaxis through the Tsr receptor. This protein is able to detect the cytoplasmic pH and periplasmic pH. It is thought that Tsr can mediate aerotaxis by detecting changes in the proton motive force of the cell.

Unlike most chemotaxis system, which have been shown to be metabolism-independent, energy taxis is metabolism-dependent.

### **1.3.6 Other factors regulating chemotaxis**

There are some factors that can activate the chemosensory system in addition to the attractants and repellents described earlier, without the need for interactions with the MCPs. It is likely that these alternative pathways connect to the MCP system at some point (95, 131). However the exact mechanism of their effects are still not clear.

#### **Intracellular $\text{Ca}^{2+}$ levels**

Increased levels of free cytoplasmic  $\text{Ca}^{2+}$  cause increased tumbling frequencies in chemotactically wild-type *E. coli*. This effect requires CheY, CheA, and CheW (212). The precise mechanism of action is unknown.  $\text{Ca}^{2+}$  may stimulate CheY or CheA phosphorylation or it may stabilize CheY-P.

#### **Intracellular fumarate level**

Fumarate has also the ability to act as a prokaryotic switch factor. Recent studies have shown that fumarate restores the ability to spontaneously switch the direction of flagellar rotation (139). No cytoplasmic constituents other than CheY are required for the fumarate effect. Malate, maleate and succinate also have the fumarate effect but to a lesser extent. Fumarate might be the connecting point between the metabolic state of *E. coli* and the chemotaxis system.

## The acetate effect

Acetate acts as a repellent that is able to cause tumbles in *E. coli*. This is called the "acetate effect". The acetate metabolism pathway uses phosphotransacetylase and acetate kinase. The intermediate molecules in this process are acetyl phosphate and acetyladenylate, shown to phosphorylate or acetylate several response-regulators, including CheY, which then brings about CW rotation (163). This phosphorylation event could therefore explain the connection between acetate and chemotaxis.

### 1.3.7 Chemotaxis in other bacteria

Chemotaxis in most prokaryotes follows the same fundamental mechanism as that used in *E. coli*. However there are also some differences between the systems, some of these will be referred to as an example.

Chemotaxis in *Bacillus subtilis* differs from that in *E. coli* in several ways. *B. subtilis* has three additional *che* genes (*cheC*, *cheD*, and *cheV*, a *cheW-cheY* fusion) but lacks *cheZ* and free CheY (7, 152). *cheZ* has only been found in enteric species. In addition, in *B. subtilis*, attractant stimulates CheA kinase, and CheY-P causes CCW flagellar rotation to mediate smooth swimming (25), which is in contrast of *E. coli* response. Both attractant and repellents stimulate hydrolysis of methyl groups in *B. subtilis* (7, 143). The role of methylation in adaptation is also different from that in *E. coli*.

Non-enteric bacteria are thought to possess more complicated sensory systems than those of enteric bacteria. Such differences may reflect the more-complex environments in which non-enteric organism live and the greater flexibility of their metabolism.

The  $\alpha$ -subgroup of Proteobacteria including *Sinorhizobium meliloti* and *Rhodobacter sphaeroides* have a motility and chemotaxis system that are different from those of enteric bacteria. In these bacteria, at least two distinct chemosensory regions are involved in chemotaxis. In *Rhodobacter sphaeroides* together, there are two copies of CheA, CheW (81, 82) and CheR, at least three CheY and one of CheB, as well as five *orfs* of unknown function and three Tlp (transducer-like protein). *R. meliloti* possesses three motility and chemotaxis regions, which contains one copy of *cheA*, *cheW*, *cheD*, *cheR* and *cheB*, two of *cheY*, one unknown *orf* and one *tlp*.

Interestingly, in both bacteria CheZ is absent but at least two copies of the response regulator CheY are present (11).

These bacteria can vary their swimming speed during chemotaxis. It has been suggested that the two CheY homologues in *Sinorhizobium* control the speed of flagellar rotation. The *S. meliloti* flagellum neither stops nor switches direction, but changes its rate of rotation. Mutation studies have been shown that CheY2 slows motor rotation in response to negative chemotactic stimulation, thereby changing the swimming direction of the cell. CheA phosphorylates the two CheYs (189). This new mechanism seems to apply to the other members of the  $\alpha$ -subgroup of Proteobacteria.

#### **1.4 Motility in *E. coli* and *S. typhimurium***

Motility and movement of bacteria is an essential complementary action for chemotaxis. Motility is more valuable outside the host and less so within. Just as with chemotaxis, the best understanding of bacterial flagella, their structure and function, is in the closely related enteric bacteria *E. coli* and *S. typhimurium*.

*E. coli* is a motile bacterium, capable of movement in a liquid medium by means of flagella. However, it should be recognized that flagellar motility is not the only form of movement in bacteria. *E. coli* has typically 6-8 flagella arising from random points on the surface of the cell. Flagellation in *E. coli* and *S. typhimurium* is random with respect to both number of flagella and sites of origin. Such cells are described as peritrichously flagellated. The pattern of flagellation also varies considerably among species. Movement of bacteria is carried out by the rotation of flagella, which can rotate in either direction, counterclockwise (CCW) causing a run or clockwise (CW) producing a tumble. Bacterial flagella are rotated by reversible motors, located at the base of each flagellum, in the cytoplasmic membrane. The transmembrane gradient of protons (protonmotive force, PMF), or in some species, sodium ions is used as the energy source for movement (21). The flagellar motor, in this aspect, is a device that converts chemical energy stored in the membrane ion gradient into the mechanical work of rotation. The mechanism of this energy conversion is not yet understood.

The flagellar motor continuously switches between CCW and CW rotation. The default direction of the flagellar motor is CCW. In counterclockwise rotation,

flagella aggregate together, forming a bundle that pushes the bacterium forward. When one or more flagella switch to CW rotation, the bundle falls apart and thus the bacteria tumble. The tumble is a result of transition of some flagella from one form to another. The motor has also been shown to pause, which is thought to be a result of incomplete switching events. For more detailed review of flagellar structure and organisation see (122).

#### 1.4.1 Flagellar Structure (21, 122)

The *E. coli* flagellum morphologically consists of a thin helical tube (20 nm in diameter), called the filament, which is joined by accessory proteins to a short curved segment named the hook. This, in turn is connected to a complex structure at the base of flagella consisting of several rings mounted on a rod, termed together the basal body. They are thought to generate flagellar rotation and act as a motor.

The filament, which is rather rigid, comprises of several thousand copies (About 20000 monomers) of a single protein called flagellin. This protein is encoded by *flgC*. The length of each filament varies and can be up to ten times the cell body length (up to 20  $\mu$ m). The long helical filament which acts as a propeller when rotate, is known to have a central channel (25-30 Å diameter) that flagellin monomers can be exported through for filament assembly (144). The tip of the filament is covered by the product of *fliD* (HAP, hook-associated protein).

The filament is connected to the cell by the hook, which is structurally similar to the filament but contains different subunits. The hook protein is composed only of *flgE* gene product. Between hook and filament are two junctions or hook accessory proteins, FlgK and FlgL (HAPs) (87). The hook is believed to act as a flexible, or universal joint between filament and cell. Hook length is controlled by a mechanism that involves FliK protein. Recently, it has been proposed that a structure in the basal body, C-ring is the element that may control the hook length, rather than FliK. There is a possibility of involvement of other proteins such as FlhB in controlling hook length. The hook also has a central channel, like the filament, allowing the export and assembly of hook and filament subunits.

The hook is connected to a structural complex known as the basal body, which is mainly located within the cell envelope. The basal body is seen to be of cylindrical symmetry, consisting of a rod joined to the hook and passing through a set of four

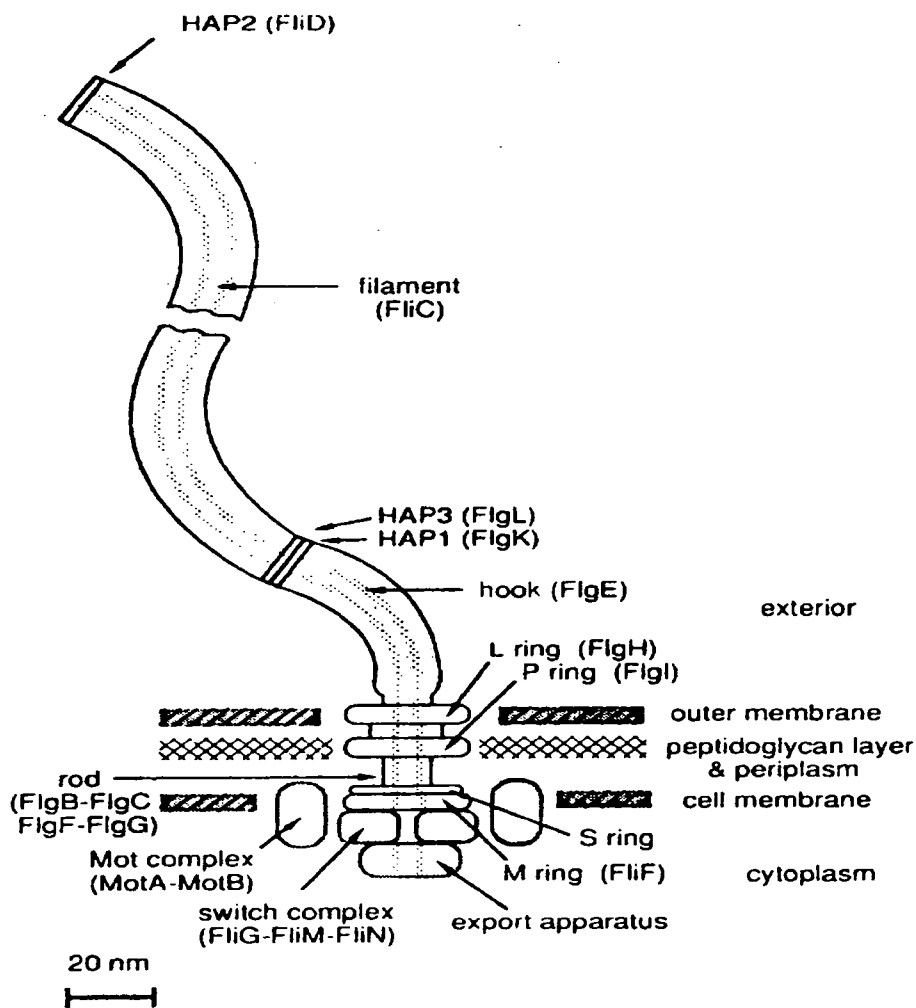
rings. The rings are named according to their position in the cell. The two outer rings are called L- and P- rings, and the inner two rings named S- and M- ring. These two rings make a double ring, called together the MS-ring, which is made up of a single protein encoded by *fliF*. The basal body is thought to function as a mechanical motor, which the MS-ring acts the rotor, the rod as the transmission shaft and the outer membrane rings are the bushing.

Recent studies showed that the rod has a helical symmetry, structure similar to that of the hook and filament, and it has the hollow central channel required for protein exportation. The channel is made from axial proteins, which form the rod, hook and filament. The rod is composed of four proteins, FlgB, C, F and G. The rod contains about 6 copies of each FlgB, C and F and about 26 copies of FlgG. The FliE protein, a component of the basal body, has also been found to be required for assembly of the rod. There are other proteins that are thought to involve in this process, causing the flagellum-specific export apparatus to function correctly. Interestingly, the proteins responsible for export and assembly of the external components of the flagella are closely related to proteins responsible for export of virulence factors (see section 1.4.3)

The L- and P- rings are made up of proteins FlgH and FlgI respectively. These two proteins form a large pore in the periplasmic space and outer membrane, through which the rod passes. The P-ring lies in the peptidoglycan layer and L-ring is embedded in the outer membrane.

The last part of the flagellum is the switch complex, which consists of a rotor and a stator. The rotor is a large ring- or bell-shaped structure, called the C-ring (C as cytoplasmic), which is attached to the cytoplasmic face of the basal body. The C-ring is composed of three switch-complex proteins, FliG, FliM and FliN, known to be important for flagellar assembly, rotation and direction control. It is thought that FliG is attached to the cytoplasmic face of the MS-ring, while FliM and FliN form the C-ring. It is still not clear whether they are a part of the rotor, or stator of the flagellar motor. FliM is exposed to the cytoplasm and is connected to both FliN and FliG. The latter connects FliM to the basal body of the flagella. The precise composition of the C-ring is unknown, but it is quite large and would contain up to 100 copies of FliM and FliN. There is a possibility that the C-ring also contains other proteins in combination with FliM and FliN. The precise roles of the three switch proteins are still unclear. FliM consists of four functional regions involving in interaction with

CheY-P, FliG, and regions used for switching and assembly of the flagella and a C-terminal required primarily for assembly (213). The amino terminus of FliM is the CheY docking site. The motor switches when CheY-P binds to FliM, generating CW



**Figure 1.4.1** Diagram showing the general structure of the *E. coli* flagellum. Figure adapted from (123). The position of the C-ring is indicated by the switch complex.

HAPs= hook-associated proteins

rotation. The processes occurring within the switch after CheY binding are not known. Some observations suggest that the switch protein recognizes the conformation of CheY, rather than its phosphoryl group. Recent evidence has revealed an interaction between FliM and FliG indicating that the binding of these two proteins might be critical for switching.

The motor consists of two parts, rotor and stator. The FliG, FliM and FliN form the rotor, whereas MotA and MotB constitute stator. FliG directly participates in rotation as it rotates along with MS-ring. FliG can be fused to FliF, which forms MS-ring. This ring is attached to the filament and believed to rotate with it, so FliG is likely to rotate also.

There are two other motor subunits involved in power production. They are MotA and MotB that seem to be important for torque generation. They form a particular ring that surrounds the MS-ring. These two proteins are membrane bound, and function in transmembrane proton conduction. MotA is a proton channel, while MotB serves as an anchor, which connects the stator or non-rotating part of the motor to the cell wall. Mutation studies strongly imply that MotA and MotB interact with each other forming a complex and the presence of MotB is necessary for MotA function. The MotA and MotB complex forms a proton-conducting channel through the bacterial cytoplasmic membrane. Each motor can contain up to 8 independent MotA-MotB complexes (26).

In most species the periplasmic domain of MotB contains a sequence motif, which could be involved in binding to peptidoglycan. Interestingly, MotB of *R. sphaeroides* lacks this sequence, exhibiting instead a heptad repeat of histidine residues (153). These sequence differences might be related to the fact that the motors of *R. sphaeroides* rotate only in one direction, stopping occasionally but never reversing. In other species, the motors alternate between CW and CCW rotation. Recent evidence suggests that the electrostatic interaction between charged residues in the stator (MotA) and the rotor (FliG) elements of the motor might play a crucial role in torque generation.

In brief, there are currently 44 known flagellar genes in *S. typhimurium* and *E. coli*. Twenty three of these encode structural components of the flagellum. Five genes are needed for torque generation. The remaining components are the filament (propellor), the hook (universal joint), and the basal body. Another five flagellar genes have regulatory roles. The remaining genes (about 11 genes) encode the



component of the flagellum-specific export that are responsible for delivery of flagellar proteins to their destination (see section 1.4.3).

#### **1.4.2 Chromosomal organization of the flagellar and chemotactic genes of *E. coli* and *S. typhimurium* (21, 26, 71, 122)**

The products of over 60 genes are required for chemotaxis, and the assembly and operation of the flagella of *E. coli* and *S. typhimurium*. There are 15-17 operons coding for flagellum proteins, whereas a few operons code for the chemosensory components. Most of them are needed for flagellar assembly, several are involved in controlling the direction of motor rotation in response to chemotactic or other sensory stimuli, and a few are involved in torque generation. The majority of these genes are highly clustered on the chromosome.

The motility and chemotaxis genes of *E. coli* are arranged in three regions and in *S. typhimurium*, they are clustered in four regions. The regions are designated as I, II, III, and IV. Region III is further subdivided into two sections, IIIa and IIIb. Each region consists of up to 14 operons, which in turn contain between 1-9 genes. The genes are clustered in operons based on their similar function, or requirement at similar step in the overall process. The genes are designated according to the region of the chromosome where they are found, however some genes are also named based on their function in the chemotaxis and motility systems. The genes in region I are designated as *flgA-M* genes, region II as *flh*, *che* and *mot*, region III as *fliA-T* and in *S. typhimurium* region IV as *flj*. The genes that are named *fli*, *flg* and *flh* are involved in the flagellar structure or assembly. The *che* genes code for proteins in the central intracellular signaling pathway of chemotaxis, while *mot* genes are part of the motor complex of the flagella.

Expression of the operons is controlled by secondary sigma factors for RNA polymerase and takes place in a hierarchical manner. One operon to be expressed requiring the expression of the lower class operon. The operons are divided into four classes: class1, 2, 3a and 3b. The lowest class in this regulatory hierarchy is class1, which contains only 1 operon, called master operon coding FlhD and FlhC. The expression of this class is required before that of any other classes. The class1 operon is under the control of cyclic-AMP (cAMP) via the regulatory protein CAP. The seven class2 operons are directly under the control of the class1 operon. Class II

operons contain a flagellum-specific consensus at –10 position of their promoters (GATAAA). *fliA*, a gene from class2 encodes a sigma factor for classes 3a and 3b. The class 3a is controlled by the master operon (class1) and *FliA*, but the class 3b is solely controlled by *fliA*. Class III promoters contain a specific cosensus at the –35 region (TAAA) in addition to the –10 cosensus sequence common to class II promoters.

### 1.4.3 Flagellar assembly and flagellum-specific export apparatus

The flagellum of *E. coli* and *S. typhimurium* is composed of four distinct parts, the motor complex, the basal body, which are embedded in the cell surface, the hook and the helical filament, both of which are external to the cell. The basal body consists of several structural subunits; called MS ring (the inner ring), P and L rings (the outer rings), the rod, and the flagellum-specific export apparatus. Biogenesis of the bacterial flagellum is thought to start from the most cell-proximal structure, the basal body, to the most cell-distal structures, the filament (122). It begins with the basal body formation. Once the MS ring is built, the rod, the P and L ring, the hook and the filament are constructed. It has been shown that flagellin monomers are added to growing filaments at their distal end (122). Modification of the flagellin subunits, or some other flagellar proteins, by glycosylation, acetylation, and methylation is required for correct assembly of the flagellin subunits into the filament. There are some experimental evidences indicating the export and usage of so-called flagellum-specific muramidase. Recent study has been shown that FlgJ protein from *S. typhimurium* has such activity. This protein is the flagellum muramidase which hydrolyses the peptidoglycan layer to assemble the rod structure in the preplasmic space (145). The function of this enzyme is to punch a hole in the peptidoglycan layer in the early stage of flagellar morphogenesis, in order to allow penetration of the nascent rod (137). Flagellum assembly also requires the function of an anti-sigma factor, FlgM and a protein that controls the length of the flagellar hook, *flgK* (71, 122, 204). FlgM binds to FliA, a specialised  $\delta$  factor, acting as an anti- $\delta$  factor that prevents the binding of RNA polymerase  $\delta^{28}$  to class III promoter. This will inhibit the expression of class III operon that contains genes necessary for filament assembly before of the hook completion. Once the hook has been assembled, the FlgM will disassociated from FliA and secreted by flagellum-specific export apparatus to the outside of the cell, resulting of FliA activity and filament assembly.

The flagellar proteins, which are synthesised in the cytoplasm, have to cross the outer membrane to reach their destinations. There are at least nine structures that are located beyond the cytoplasmic membrane, and their proteins should be delivered to their locations. It has been proposed that the central channel existing within the growing flagellar structure is the physical pathway of protein transportation (107, 121, 122). It is thought that the proteins are exported to the periplasm, the outer membrane, or the extracellular space, where their assembly finally occurs by two distinct export mechanisms. One, which is more common, called type II *sec*-dependent pathway or GSP, is used for export of proteins with cleavable signal sequences (121). Only the P and L rings (FlgI and FlgH, respectively) have been found to be synthesised as proteins with cleavable sequences (161), and hence are transported across the cytoplasmic membrane, to the periplasm by the protein export system. The other mechanism, which is unusual, is called *Sec*-independent export pathway or typeIII flagellum-specific export pathway/apparatus (71). The distinct feature of this export pathway is that it transports proteins, which lack a cleavable signal sequence necessary for transportation with the general pathway. Twelve proteins of the filament, rod, hook, hook-associated proteins, basal body (FlgB, FlgC, FlgF, flgG, flgE, FlgK, FlgL, FliD, FliE and FliC), and FlgM, the regulatory protein (anti-sigma factor) do not carry amino-terminal leader peptides signal and are thus exported from the cytoplasm to their final destination via the flagellum-specific export pathway. Since most components of the flagellum are outside the cytoplasm and do not have signal sequences, the function of the export apparatus plays an important role in flagellar assembly. It has been shown that this system contains at least eight general components, which could rise to 13 if the genes with specific functions such as chaperones are included. These eight general components, identified by analysis of various conditional and knockout mutants, are *flhA*, *flhB*, *fliH*, *fliI*, *fliO*, *fliP*, *fliQ* and *fliR* without which no export of any kind can occur. All components of the flagellum-specific export pathway are conserved among different bacteria. Two proteins, FliJ, a cytoplasmic protein and FliE, a basal body component, were found to be necessary for export of at least some substrates such as FlgD (hook-capping protein) and FlgE (hook protein) (122, 137). FliJ seems to act as a general component of the typeIII flagellar export apparatus, rather as a specific Chaperone (138). It is believed that in a short period of time, the L ring (FlgI protein) alone or as a complex L ring-FlgH is also required for export of proteins, since a temperature-sensitive *fliN* mutant cannot

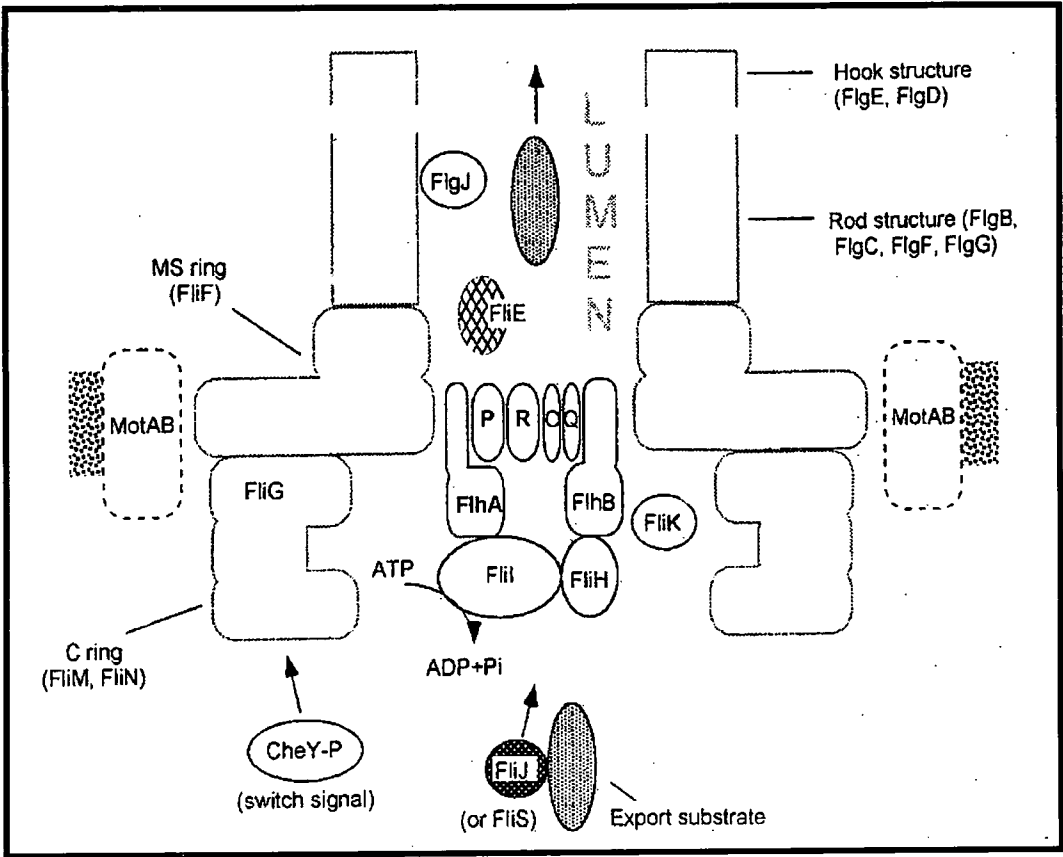
export flagellar proteins (138). FlhA, FlhB, FliO, FliP, FliQ and FliR are all integral membrane proteins, three of which are physically associated with the flagellar basal body. The proposed location of the membrane-associated components is in a central pore within the MS ring of the flagellar basal body. There is physical evidence that such a pore exists (138) and it has been shown that FlhA, FliP and FliR are associated with the basal body (137). The MS ring and C rings are also required for export of all proteins as they provide a physical support structure for the export apparatus, but they are not directly involved in export. The biochemical functions of most of these genes are not yet known. One member of this system, FliI, was found to have sequence similarity to subunits of the proton-translocating FoF<sub>1</sub> ATPase, and is believed to supply the energy necessary for export.

It has been suggested that export proteins, with the aid of cytosolic chaperones such as FliJ or FliS, are brought into contact with two soluble components, FliI (an ATPase) and FliH (function unknown) to form FliH-FliI-FliJ-protein complex (138). They deliver the proteins in an energy-dependent process to the soluble domains of FlhA and FlhB and release their substrates. FlhA and FlhB have large soluble domains that protrude into the cytoplasm within the cavity that lies at the centre of the cytoplasmic C ring of the motor. The substrates are then translocated across the plane of the membrane through a complex of FliO, FliP, FliQ, FliR and the transmembrane domains of FlhA and FlhB. At the end the translocated substrates cross the channel and assemble at its destination. The export apparatus will not export filament-type proteins until it has received a signal to switch its specificity, and this signal is generated only upon completion of the hook. The switch in export specificity is mediated by FlhB, by blocking the gate for the hook protein, but opening it for the filament protein, along with FliK that controls the hook-length.

One regulatory feature, FlgM, an anti-sigma factor which binds to FliA to prevent its association with RNA polymerase enzyme (150), is exported from the cell by the typeIII flagellum export apparatus.

Members of the flagellum-specific export pathway have similarity to components of a newly identified protein export system called the typeIII secretion pathway or contact-dependent system, which is responsible of the export of a variety of virulence factors, by pathogenic bacteria, into plant and animal host cells. For instance, FliI was found to have homology to Spa47 of *Shigella flexneri* and HrB6 of *X. campestris*, FlhA to MxiA of *S. flexneri*, InvA of *S. typhimurium* and LcrD of

*Yersinia pestis*, FliR to MopE of *Erwinia cartovora* and SpaR (Spa29) of *S. flexneri*. The typeIII secretion system is a protein delivery machinery that enables bacteria to translocate pathogenicity proteins from the cytoplasm across the inner and outer membrane, into the plant and animal cells. It is composed of approximately 20 proteins, most of which are located in the inner membrane and build a large structure that spans both bacterial membrane and possibly the host cell membrane as well. The homology between members of two systems should not be surprising, since both systems are responsible for transportation and localization of proteins across the cell



**Figure 1.4.3.1 Diagram showing Hypothetical structure of typeIII Flagellum-specific export apparatus.** Figure adapted from (138)

membrane. It has become clear in recent years that the flagellum-specific export apparatus and the typeIII secretion system have common characteristics (92, 138). The two systems constitute a superfamily of proteins that share similar overall structural features, but are involved in various cellular processes including virulence in *Yersinia pestis*, host invasion in *S. typhimurium* and *Shigella flexneri*, and flagellar assembly in *E. coli* and *C. crescentus*. The proteins exhibit several putative transmembrane domains in the amino-terminals, suggesting that they are membrane integral proteins. The proteins are exported by these specific pathways also share several characteristics. For example, harpin (Hrp), Yop, Ipa, Inv and flagellar proteins are all cell surface-associated and extracellular components. In addition, they do not contain the conserved N-terminal signal sequence characteristic of proteins secreted by the *Sec*-dependent pathway of Gram-negative bacteria. It is thought that a typeIII secretion signals is located at the amino terminals of the both flagellar and virulence proteins (19). It has been proposed that such a signal resides in the 5' end of the protein mRNA (6). A second secretion signal has been described in YopE which is dependent on interaction with its chaperones (19). The conservation between genes of typeIII secretion machines and flagellar-specific export complex compelling and speculate that latter may also recognize mRNA signal to export flagellar units across the bacterial membrane (5).

Some lines of evidence suggest that the virulence export pathway may also be used for export of flagellar components, since a *mopE* mutant, a *fliR* homologue in the typeIII secretion system which is responsible for export of virulence factors in *Erwinia carotovora*, failed to assemble the flagella, in addition to a reduction in virulence level in plants. In the typeIII secretion system, a large homomultimeric complex, made of molecules called secretins, is found in the outer membrane. It is thought that the function of this complex is to act as a passage for export of substrates across the outer membrane. A needle-like structure with a hollow projection extending out from the bacterial surface, that contains a known component of the typeIII export system, and closely resembles a flagellar basal body, was recently found in *Salmonella* (106). This needle complex structure may play the same role as the basal body, in the typeIII secretion system. It has been postulated that the L-ring of the flagellar basal body may transiently correspond to a secretin-like structure that links the two systems of export of flagella and virulence proteins (137). It has recently been revealed that flagellar export machinery, like typeIII system has chaperones. The role

of chaperones is directly interaction with secreted proteins and is responsible for preventing folding or aggregation. Chaperones might have dual roles in the cytoplasm. They may preventing premature association of their target substrates, and also involved in substrate recognition by the export systems in the membranes (19, 119). FlgN, FliT and FliS have been found to facilitate export of flagellar proteins such as FliC and have chaperone-like functions. These evidences are strengthening the idea that typeIII flagellar export system and typeIII virulence secretion system closely related. In fact, the flagellar export apparatus is a sort of typeIII secretion pathway, differing only in the nature of its export proteins. It seems possible that the origin of the typeIII secretion system may have come from the flagellar protein export system, since flagella are very ancient organelles. Perhaps in evolution the flagella-specific export pathway has been changed for the elaboration of some pathogenicity factors in bacteria. For more review about relation between flagellar- specific export pathway and typeIII see (19, 92, 119). Interestingly, some function and structural similarities have also been found between the typeIV DNA-transfer system which is responsible to deliver *Agrobacterium* T-DNA into host cells, and typeIII secretion system (46). VirB10 and VirB11 are homologues to members of various bacterial systems for transport of virulence proteins (89, 239, 240). Homologies were also reported between VirB4 and FliI proteins from flagellar export system as both are ATPase. These evidences suggest that T-DNA transport system may have evolved from a protein transport system.

Many components of the flagellum-specific export pathway have also been found in *A. tumefaciens* and the closely related bacterium *Sinorhizobium meliloti*. These genes have sequence similarity to members of the system in many other gram-negative and Gram-positive bacteria, and different genes of typeIII secretion systems of Gram-negative bacteria. However, the order of the export genes in *A. tumefaciens* and *S. meliloti* is different from those in *E. coli*, *C. crescentus* and *B. subtilis*. Surprisingly, genes controlling cultivar-specific nodulation of soybean by *S. fredii* have been found to have sequence similarity to the genes encoding the typeIII secretion system in other bacteria. Homology between symbiotic genes and typeIII secretion proteins (73), and similarity between the flagellum-specific export pathway and the typeIII system, reinforces the idea that these three systems, symbiosis, pathogenesis and flagellar assembly, share common molecular mechanisms.

#### 1.4.4 Motility and Chemotaxis in *Agrobacterium tumefaciens*

*Agrobacterium tumefaciens* is a motile bacterium. It can recognise some compounds released from wounded cells, and move to the source of these chemicals (13, 191, 192). Motility of the bacterium is carried out by bacterial flagella. Electron microscopy has revealed that *A. tumefaciens* has an unusual flagellation pattern. The bacterium typically has a polar tuft of two flagella, plus two to four lateral flagella with curved filaments, which are composed of up to four different flagellins. The flagella, which are generally longer than the cell itself, only rotate in one direction, clockwise (CW), producing a run. The length of run is relatively long (frequently in excess of 500µm in C58C1 strain), and is rarely interrupted by changes of direction such as those seen in *E. coli* (117, 118, 180). The mechanism used by *A. tumefaciens* to change direction is unknown. These behavioural features of *A. tumefaciens* are similar to *Sinorhizobium meliloti*, but are different from those of *E. coli*. Motility and chemosensory system of the  $\alpha$ -subgroup of proteobacteria, such as *Agrobacterium*, *Azospirillum*, *Caulobacter* and *Rhizobium* are distinctly different from the enterobacterial paradigm (8). Some differences has been described in section 1.3.7. One distinguishing features between two groups is existence of two different CheY acting with new MotC and MotD motor proteins to control flagellar rotatory speed (79, 189). *Agrobacterium* flagella structure are also different from *E. coli*. Filaments of *E. coli* possesses only one protein (FliC) whereas *A. tumefaciens* filaments is composed of several types of flagellin subunits. Four flagellins contribute to *Agrobacterium* filament structure: FlaA is the major protein, FlaB and FlaC are present in lesser amounts, and FlaD is minor component.

The strain C58C<sup>1</sup> used in this work, has been shown to be an active swimmer, with long straight or curved runs. Other *A. tumefaciens* genotypes such as LBA4301 and A136 are relatively poorly motile in comparison to C58C<sup>1</sup> (14, 118).

It has been suggested that most, if not all, motile bacteria use the MCP chemotactic signaling system found in enteric bacteria. A number of experiments indicated that *A. tumefaciens* also possesses MCP-like proteins and employ a similar signal transduction pathway to *E. coli*. When *A. tumefaciens* is exposed to a gradient of attractant, it will migrate in a favourable direction (14). The involvement of methylation for chemotaxis of *A. tumefaciens* to sucrose was demonstrated by Loake



(118). Experiment carried out by Brown *et al.* showed that *Agrobacterium* requires methionine for chemotaxis to occur (34). An open reading frame has also been identified by Farrand and his co-workers which shows significant identity to a *C. crescentus* MCP (70). Another experiment showed that some *Agrobacterium* DNA fragments hybridised with a synthetic oligonucleotide complimentary to the conserved signaling domain of *E. coli* MCPs (149). VirA, a member of the two-component virulence system, structurally resembles MCPs and has sequence conservation with CheA. VirG, another member of the same two-component system also possesses similarity to CheB and CheY. A report showing that a protein with significant homology to a known MCP has also been found in the Ti plasmid pTiC58. Taken together with other results described earlier, this evidence indicates the existence of methylation-dependent chemotaxis in *A. tumefaciens*.

*Agrobacterium* has been shown to possess a highly sensitive chemotaxis system, with threshold responses occurring below  $10^{-8}$  M (181). However, in most cases the *A. tumefaciens* response is more sensitive than *E. coli* (118). This chemotactic interaction with possible host plants of *Agrobacterium* appears to involve two distinct signaling pathways. A general chromosomally encoded chemotaxis pathway, which responds to nutrients such as amino acids and sugars, is responsible for maintaining the bacteria in sufficiently large number in the rhizosphere for infection to take place (118). Another system is a plasmid encoded pathway, which is responsible for movement of the bacterium to susceptible plant cells (12).

### **1.5. Analysis of chemotaxis and Motility genes in *Agrobacterium tumefaciens***

In order to study *Agrobacterium* motility and chemotaxis genes in more detail, a number of chemotactic mutants were generated using a Tn5 suicide vector, followed by selection for non-chemotactic mutants (117, 180). Further experiments associated with complementation and hybridization showed that similar to *E. coli* and other motile bacteria, the motility and chemotaxis genes in *A. tumefaciens* were thought to clustered together on the chromosome. A number of genes involved in flagellar structure, assembly, function and switching have been found on cosmid pDUB1900 from an *A. tumefaciens* genomic library. Analysis of 22 kb of sequence from this cosmid have identified an operon encoding homologues to the flagellar switch

proteins FliM, FliN and FliG (56), a group of flagellar genes consisting of *flgB*, *flgC*, *flgG*, *fliE*, *flgI* and *flgH* whose protein products are involved in basal body formation, a possible flagellum export gene (*fliP*) and two genes of unknown function (57), *flgA* and *fliL*. FlgBCG and E are rod proteins, whereas FlgI and H are L and P ring proteins. Recently four flagellin genes, *flaA*, *B*, *C*, and *D* which are concerned with flagellar filament structure have been found in this cluster (55-57). FlaA is the major protein, FlaB and FlaC are present in lesser amount, and FlaD is a minor component. An additional putative operon including the homologues to the *fliI*, *flhB* and *flgF* were also found in this region.

A cluster of chemotaxis genes has been identified on cosmid pDUB1911 from a representative genomic library of C58C<sup>1</sup> (228). The cluster consists of *orf1*, whose predicted protein shows strong homology to an MCP, followed by *orf2*, *cheY1*, *cheA*, *cheR*, *cheB*, *cheY2*, *orf9* and *orf10*. A homologue to the flagellar gene *fliF* was found directly downstream of this cluster.

## 1.6 Aims of this project

The main goal of this work was to further characterise the motility and chemotaxis genes in *A. tumefaciens* using cosmid pDUB1905. This cosmid had previously been shown to contain *che-2* sequence and sequences spanning the site of two other mutations, *fla-8* and *mot-6* (34, 180). These mutants had been created by insertion of the transposon Tn5. Electron micrographs and swarming experiments of these three mutants showed that the mutant *che-2* is a non-chemotactic flagellated motile mutation, which tumbles excessively. The non-motile mutants *fla-8* and *mot-6* were originally classified separately as by electron microscopy, *fla-8* did not possess flagella, whereas *mot-6* possessed normal flagella. The two latter mutants were shown to share a common site in the cosmid pDUB1905, their apparent phenotypes probably resulting from an unexpected error. Finding the real phenotype of *fla-8* and *mot-6* mutants was another goal of this project.

The region of pDUB1905, spanning the mutant sequences was initially to be sequenced and analysed. Depending upon the result, any flagellar or chemotaxis gene homologues identified, were to be further characterised. Functional studies of each gene were to be achieved using site-direct mutagenesis by insertion of a neomycin

cassette. This was to be followed by phenotypical analysis of the resulting mutant strains with light and electron microscopy, and in the swarm plate assays.

Preliminary evidence showed that the chemotaxis and motility genes are clustered together, in separate chromosomal regions (34, 180). The result of an early experiment suggested that the two cosmids, pDUB1900 and pDUB1905, which contain several chemotaxis and motility genes might be close to each other (55). Finding how close these two fragments was another goal of this project.

## **2. Materials and Methods.**

## 2.1 Materials.

All inorganic chemicals were of AnalaR quality and purchased from BDH Chemicals Ltd., Poole, Dorset, U.K. unless otherwise specified.

All organic chemicals and enzymes were from Sigma Chemicals Plc., Poole, Dorset, U.K. unless otherwise specified.

Lab M Nutrient Broth (no.2), Lab M Nutrient Agar, nylon hybridisation transfer membranes and radiochemicals were from Amersham Ltd., Bury, U.K..

Agar bacteriological (no.1) and yeast extract were from Oxoid Ltd., Basingstoke, Hants., U.K.

Restriction endonucleases, T4 DNA ligase, Klenow enzyme, Taq polymerase, corresponding buffers, DNA markers, X-gal, IPTG and wild type  $\lambda$  DNA were from NBL, Cramlington, Northumberland, U.K., Boehringer Mannheim (UK) Ltd., Lewes, U.K., or Helena Biosciences, Tyne & Wear, U.K..

Agarose was from BRL, Gaithersburg, U.S.A.

Ficoll 400 and Sephadex G-50 were from Pharmacia Fine Chemicals, Uppsala, Sweden.

Fuji RX-100 X-ray film was from Fuji Photo Film Co., Ltd., Japan.

Polaroid film was from Polaroid (UK) Ltd., St. Albans, Hertfordshire, U.K..

Filter paper (3MM) and laboratory sealing film were from Whatman International Ltd., Maidstone, U.K..

Minisart filters were from Sartorius GmbH, Postfach 3243, D-3400 Göttingen, Germany.

Nitrocellulose discs (25 mm, 0.22 µm pore size) for tri-parental matings were from Schleicher and Schuell, Postfach 4, D-3354, Dassel, Germany.

Oligonucleotides for use in sequencing and PCR analysis were from MWG-Biotech, Germany and PE-Applied Biosystems UK, Cheshire, U.K.

2.2 Bacterial Strains and Plasmids.

2.2.1 *E. coli* Strains.

Strain	Characteristics	Resistance	Source
DH5α	F- <i>supE44 ΔlacU169 (f80 lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1</i> .		Lab. Stock (227)

2.2.2 *A. tumefaciens* Strains.

Strain	Characteristics	Resistance	Source
C58C <sup>1</sup>	Ti-cured strain of C58.	Rif <sup>R</sup>	Lab. Stock (217)
C58C <sup>1</sup> / <i>fliR</i>	Behavioural mutant - created by insertion of a neomycin resistance cassette into <i>fliR</i> .	Rif <sup>R</sup> , Neo <sup>R</sup>	This study

2.2.3 Plasmids.

2.2.3.1 Plasmid Vectors.

Plasmid	Characteristics	Resistance	Source
---------	-----------------	------------	--------

<b>pBluescriptII SK<sup>+</sup></b>	2.96kb <i>E. coli</i> vector.	Amp <sup>R</sup>	Lab. Stock
<b>pJQ200mp18</b>	Suicide vector for <i>A. tumefaciens</i> with polylinker of M13mp18 and <i>sacB</i> .	Gm <sup>R</sup>	Lab. Stock (162)
<b>pJQ200SK<sup>+</sup></b>	Suicide vector for <i>A. tumefaciens</i> with polylinker of SK <sup>+</sup> and <i>sacB</i> .	Gm <sup>R</sup>	Lab. stock (162)
<b>pJQ200uc1</b>	Suicide vector for <i>A. tumefaciens</i> with <i>sacB</i> .	Gm <sup>R</sup>	Lab. stock

2.2.3.2 Plasmids Used in Mutant Construction.

Plasmid	Characteristics	Resistance	Source
<b>pRK2013</b>	Helper plasmid for conjugation in <i>A. tumefaciens</i> .	Kan <sup>R</sup>	Lab. stock
<b>pDUB2033</b>	Neomycin resistance cassette.	Amp <sup>R</sup> Neo <sup>R</sup>	Lab. stock

2.2.3.3 Recombinant Plasmid Containing Cloned behavioural Genes.

Cosmid	Characteristics	Resistance	Source
<b>pDUB1905</b>	A C58C1 cosmid library clone containing a <i>Bam</i> HI chromosomal fragment in the vector pLAFR-3 (carries behavioural genes).	Tc <sup>R</sup>	Lab. Stock (180)

**2.2.3.4 Recombinant SK+ Plasmids Containing Subcloned *Agrobacterium* Behavioural Genes.**

Plasmid	Characteristics	Resistance	Source
pAB1	pBR322 containing 12kb <i>Bam</i> HI fragment from pDUB1905.	Amp <sup>R</sup>	Lab. Stock (34)
ph1	pBluescriptII SK <sup>+</sup> containing 5.0kb <i>Eco</i> RI fragment from pDUB1905.	Amp <sup>R</sup>	This study
ph2	pBluescriptII SK <sup>+</sup> containing 1.7kb <i>Eco</i> RI- <i>Bam</i> HI fragment from pDUB1905.	Amp <sup>R</sup>	This study
ph3	pBluescriptII SK <sup>+</sup> containing ~3kb <i>Eco</i> RI fragment from pDUB1905.	Amp <sup>R</sup>	This study
ph4	pBluescriptII SK <sup>+</sup> containing 0.3kb PCR-synthesised fragment from chromosomal DNA, adjacent to pDUB1905.	Amp <sup>R</sup>	This study
ph1.1	pBluescriptII SK <sup>+</sup> containing 3.7kb <i>Eco</i> RI- <i>Hap</i> I fragment from pDUB1905.	Amp <sup>R</sup>	This study
ph1.11	pBluescriptII SK <sup>+</sup> containing 1.7kb <i>Eco</i> RI- <i>Eco</i> RV fragment from pDUB1905.	Amp <sup>R</sup>	This study
ph1.12	pBluescriptII SK <sup>+</sup> containing 1.5kb <i>Eco</i> RV fragment from pDUB1905.	Amp <sup>R</sup>	This study
ph1.13	pBluescriptII SK <sup>+</sup> containing 0.8kb <i>Eco</i> RV fragment from pDUB1905.	Amp <sup>R</sup>	This study

**2.3 Bacterial Growth Media, Conditions and Procedures.**

The following media were used in this work:

Lab M nutrient broth no.2 (LM-broth):

25g made up to 1 litre with distilled water gives final concentrations of 10g.l<sup>-1</sup> beef extract, 10g.l<sup>-1</sup> balanced peptone no.1, 5g.l<sup>-1</sup> NaCl, pH 7.5 ± 0.2.



For swarm plates bacteriological agar was added at a concentration of 0.15 - 0.2%.

Lab M nutrient agar (LM-agar):

28g made up to 1 litre with distilled water, gives final concentrations of 5g.l<sup>-1</sup> peptone, 3g.l<sup>-1</sup> beef extract, 8g.l<sup>-1</sup> NaCl, 12g.l<sup>-1</sup> agar no.2, pH 7.3 ± 0.2.

Antibiotics were added to media after autoclaving, to the following concentrations:

For *A. tumefaciens*; chloramphenicol 25µg.ml<sup>-1</sup>, kanamycin 50µgml<sup>-1</sup>, neomycin 100µg.ml<sup>-1</sup>, rifampicin 100µg.ml<sup>-1</sup>, streptomycin 300µg.ml<sup>-1</sup> and tetracycline 15µg.ml<sup>-1</sup>.

For *E. coli*; ampicillin 50µg.ml<sup>-1</sup>, chloramphenicol 34µg.ml<sup>-1</sup>, gentamycin 10µg.ml<sup>-1</sup>, kanamycin 50µg.ml<sup>-1</sup>, neomycin 100µg.ml<sup>-1</sup>, streptomycin 20µg.ml<sup>-1</sup>, and tetracycline 15µg.ml<sup>-1</sup>.

When selecting for the inactivation of the β-galactosidase gene, by insertion of DNA fragments into the multiple cloning sites of pBluescriptII and other plasmids, 40µl of 20mg.ml<sup>-1</sup> X-gal (in DMF) were spread over the surface of agar plates.

Liquid bacterial cultures were incubated on an orbital shaker at 200rpm at temperatures of 37°C for *E. coli* and 28°C for *Agrobacterium* strain. Short-term (1-2 months) stocks of cultures were kept at 4°C on solid agar plates. Long-term stocks were kept in 40% glycerol at -80°C.

Liquid cultures were inoculated with a flamed loop or a sterile cocktail stick. Solutions and bacterial cultures were spread onto agar plates using a glass spreader which had been sterilised in 70% ethanol. Bacterial colonies were inoculated into the centre of swarm plates using a sterile needle.

Aseptic technique was used throughout with bacterial cultures. All glassware, plasticware and other equipment were sterilised by autoclaving at 121 °C, 15 psi. for 15 minutes. Generally solutions were prepared according to Sambrook *et al.*(169) and autoclaved, as above, if possible. Otherwise solutions were filter-sterilised through a 0.22 µm nitrocellulose filter into a sterile container.

## **2.4 Isolation of DNA.**

### **2.4.1 Alkaline Lysis Plasmid Minipreps.**

This method, used to prepare small amounts of relatively pure plasmid DNA, was according to Sambrook *et al.*(169)

A single colony of bacteria was grown overnight in 5ml of LM-broth containing the appropriate antibiotic selection. 1.5ml of this culture was pipetted into a sterile eppendorf tube, and the cells harvested by centrifugation for 1 minute in a microfuge. The supernatant was discarded, and a further 1.5ml of the culture pelleted as before. The supernatant was again discarded. The pellet was resuspended in 200µl of ice-cold solution 1 (1% glucose, 10mM EDTA pH8.0, 25mM Tris.HCl pH8.0), 200µl of solution 2 (0.2M NaOH, 1% SDS) was then added and the contents of the tube mixed by gentle inversion. 200µl of ice-cold solution 3 was added to the mixture and the tube inverted ten times to mix the contents. (Solution 3 was prepared by adding 1.15ml of glacial acetic acid to 2.85ml of distilled water and then adding 6ml of 5M potassium acetate). The solution has an overall pH of 4.8 and is 3M wrt potassium and 5M wrt acetate.) The tube was then microfuged for 5 minutes to remove bacterial debris. The supernatant was transferred to a fresh tube, spun down for a further 2 minutes and then pipetted into another fresh tube. The DNA was precipitated by the addition of 1ml of 100% ethanol. The tube was left for 5-15 minutes at -20°C then the DNA collected by centrifugation for 5 minutes. The DNA pellet was washed in 1ml 70% ethanol and air dried for 20-30 minutes. The final pellet was resuspended in 50µl of TE buffer and RNAase A added to a concentration of 20µg.ml<sup>-1</sup>

When plasmid DNA was to be prepared for sequencing, a QIAGEN QIAprep spin plasmid miniprep kit was used according to the manufacturers instructions.

#### **2.4.2 Larger Scale Plasmid Preparation.**

Single colonies were inoculated into two 50ml of LM-broth with antibiotic selection and grown overnight with shaking. The cultures were transferred to two 50ml centrifuge bottles and the bacterial cells harvested by centrifugation at 7000g/4°C for 7 minutes in a Beckman J2-HS centrifuge, using a JA-20 rotor. The supernatants were removed and the pellets resuspended in 3.335ml of ice-cold solution 1. 3.335ml of solution 2 was added to each tube and the tubes gently mixed. 3.335ml of ice-cold solution 3 was then added to each tube, and the tubes mixed well. The tubes were centrifuged for 5 minutes, at 4°C and 15000g in a Beckman J2-HS centrifuge, to remove cell debris. The supernatants were transferred to sterile 50ml Falcon tubes, 20ml of phenol:chloroform:isoamyl alcohol (25:24:1) were added and the tubes mixed well. The tubes were spun at 3500g for 3 minutes and the aqueous layer transferred to sterile centrifuge tubes. 20ml of ice-cold 100% ethanol was added to each tube and the tubes left on ice for 25 minutes. The DNA was pelleted by centrifugation for 15 minutes, at 4°C and 15000g in a Beckman J2-HS centrifuge. The pellet was then washed in 70% ethanol, dried and resuspended in 500µl of TE.

Solutions 1, 2 and 3 were made as in section 2.4.1

#### **2.4.3 Small Scale Preparation of Bacterial Chromosomal DNA.**

A 5ml LM-broth culture of bacteria was grown to stationary phase and 1.5ml of the culture transferred to an eppendorf tube. The cells were spun down for 1 minute in a microfuge, the supernatant removed and the pellet left at -20°C for 30 minutes. The pellet was then resuspended in 200µl of TE, and the cell suspension incubated with 8µl of lysozyme (10mg.ml<sup>-1</sup> stock) for 30 minutes at 37°C. The cells were then lysed by the addition of 40µl of 4M sodium perchlorate, 24µl of 10% SDS and 8µl of Proteinase K (20mg.ml<sup>-1</sup> stock), and incubated at 45°C for 2 hours. The DNA was precipitated by the addition of 2 volumes of ethanol and incubation at -20°C for 30 minutes. The DNA was pelleted by centrifugation in a microfuge for 5 minutes. The pellet was washed with 70% ethanol, dried and resuspended in 500µl of TE buffer.

The DNA solution was extracted three times with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1), the aqueous layer being transferred to a new eppendorf tube each time. The DNA was then extracted once with chloroform:isoamyl alcohol (24:1). The DNA in the final aqueous layer was precipitated by the addition of 2 volumes of ethanol and 0.1 volumes of 3M sodium acetate (pH 4.8), followed by incubation overnight at -20°C. The DNA was pelleted by centrifugation in a microfuge for 5 minutes. The pellet was then washed with 1ml of 70% ethanol, dried and resuspended in 50µl of TE buffer with RNAase A (10mg.ml<sup>-1</sup> stock) added to a final concentration of 20µg.ml<sup>-1</sup>.

#### **2.4.4 Small Scale Preparation of Bacterial Chromosomal DNA - 2.**

This method was used as a quick preparation of *Agrobacterium* DNA when screening for possible mutant strains.

A 5ml LM-broth culture of bacteria was grown to stationary phase and 1.5ml of this culture transferred to an eppendorf tube. The cells were spun down for 1 minute in a microfuge, the supernatant removed and the pellet resuspended in 200µl of lysis buffer ( lysis buffer consists of 40mM Tris-acetate pH7.8, 20mM Sodium acetate, 1mM EDTA and 1% SDS ). 82.5µl of 4M NaCl was then added and the tube contents mixed by inversion. The tubes were centrifuged for 10 minutes in a microfuge. The supernatant was then pipetted into a clean eppendorf tube and an equal volume (approximately 250µl) of chloroform:isoamyl alcohol (24:1) added. The tube was inverted 50 times, then centrifuged for 3 minutes in a microfuge. The upper aqueous layer was removed to a clean eppendorf and the DNA precipitated by the addition of two volumes of 100% EtOH and incubation at -20°C for 10 minutes. The DNA was pelleted by centrifugation in a microfuge for 5 minutes. The pellet was then washed with 1ml 70% ethanol, dried and resuspended in 30µl of TE buffer with RNAase A (10mg.ml<sup>-1</sup> stock) added to a final concentration of 20µg.ml<sup>-1</sup>.

### **2.5 DNA Manipulations.**

#### **2.5.1 Phenol:Chloroform Extraction of DNA.**

A 25:24:1 solution of phenol:chloroform:isoamyl alcohol was equilibrated 3 times with TE buffer and stored, under TE, in a light-proof bottle at 4°C. To remove proteins from DNA solutions an equal volume of phenol:chloroform:isoamyl alcohol was added, the solutions mixed by vortexing for 30 seconds and the phases separated by centrifugation for 2 minutes in a microfuge. The aqueous phase was transferred to a fresh tube. This was repeated until no further protein was visible (as a white precipitate) at the boundary of the two phases.

### **2.5.2 Restriction Endonuclease Digestions.**

Digestions were carried out according to the enzyme manufacturer's instructions. Generally plasmid DNA was digested in a total volume of 10-40µl, with 5 units of restriction endonuclease, 0.1 volumes of the supplied 10x concentrated enzyme buffer, and sterile distilled water to make up the volume. The reaction was incubated at the recommended temperature (usually 37°C) for 1-3 hours. If more than one restriction enzyme was to be used in the same reaction and the buffers supplied differed, the reaction was buffered using one-phor-all buffer PLUS (Pharmacia). Chromosomal DNA was digested in a larger volume, 100-200µl, with 10 units of restriction enzyme added for every microgram of DNA, as well as the appropriate amounts of buffer and sterile distilled water. The reaction mixture was covered with a layer of mineral oil to prevent evaporation and maintain the buffering conditions, before being incubated overnight at the required temperature.

If the digestions were to be analysed by gel electrophoresis, 0.2 volumes of 6x gel-loading buffer were added (6x gel-loading buffer contains 0.25% bromophenol blue, 0.25% xylene cyanol FF and 40% sucrose in distilled water - this was filter sterilised and stored at 4°C).

If the digested DNA was to be used in further subcloning steps, the required fragment was cut out of the gel and purified using silica fines (see section 2.5.4).

### **2.5.3 Agarose Gel Electrophoresis.**

Gel electrophoresis of DNA samples was carried out with large 180x200mm maxigels (volume 400ml), 100x80mm minigels (volume 50ml) or 77x55mm minigels (volume 35ml). Maxigels were run in a Pharmacia gel apparatus GNA-200

electrophoresis tank. Minigels were run in Pharmacia gel apparatus GNA-100 electrophoresis tanks. The concentration of agarose within a gel could be varied depending on the size of DNA to be separated. Usually a 0.7% agarose gel was used, which efficiently separated linear DNA between 10-0.8kb. The required amounts of agarose and 1x TAE buffer (50x stock - 242g Tris, 100ml EDTA pH8.0, 57.1ml glacial acetic acid per litre) were mixed and the agarose dissolved by microwaving the mixture. The solution was cooled to about 60°C, 10mg.ml<sup>-1</sup> ethidium bromide was added to a final concentration of 0.2µg.ml<sup>-1</sup>, and the agarose was poured into a gel mould with a well comb in place. Once the agarose had set the gel was put in a tank and covered with 1x TAE buffer containing 0.2µg.ml<sup>-1</sup> ethidium bromide. The DNA samples (and size markers) were loaded and electrophoresis carried out at 0.1A (5-10 V.cm<sup>-1</sup>) for the required amount of time, usually 1 - 1.5 hours. The size markers used were either λ-DNA digested with *Pst*I and/or λ-DNA digested with *Hind*III.

When DNA samples less than 1kb in size were to be analysed by agarose gel electrophoresis, Metaphor agarose and TBE buffer (5x stock - 54g Tris, 27.5g Boric acid, 20ml 500mM Na<sub>2</sub>EDTA (pH 8.0) made up to 1 litre with dH<sub>2</sub>O) were used. Usually 2% Metaphor agarose gels were made with 1x TBE, using the same methodology as above.

***Pst*I digested λ-DNA produces DNA fragments of the following sizes (in kb):**

14.05, 11.49, 5.07, 4.75, 4.51, 2.84, [2.56, 2.46, 2.44], 2.14, 1.99, 1.70, 1.16, 1.09, 0.81, 0.52, 0.47, 0.45, 0.34...

The fragments enclosed in brackets run together on an agarose gel. Smaller fragments are also produced, but were rarely seen in this work.

***Hind*III digested λ-DNA produces DNA fragments of the following sizes (in kb):**

23.13, 9.42, 6.56, 4.36, 2.32, 2.03, 0.56, 0.13.

DNA within the gel was visualised on a transilluminator (UVP Inc.), and photographed with a Polaroid RP4 Land camera (using a red filter) onto Polaroid 667 film.

#### **2.5.4 Isolation of DNA Fragments from Agarose Gels Using Silica Fines.**

The required band was cut out of the gel using a sterile scalpel blade, and placed in a labelled 1.5ml eppendorf tube. 800µl of sodium iodide solution were added to the tube and the tube placed at 70°C for 5 minutes to melt the agarose. Once the agarose had completely melted the tube was mixed by inversion and allowed to cool to room temperature for 5 minutes. 5µl of silica fines were then added, mixed and the solution left for 10 minutes at room temperature with continuous shaking. The fines were spun down for 30 seconds in a microfuge, the supernatant removed and the fines washed with 70% ethanol. The silica fines pellet was dried with a piece of tissue, then resuspended in 30µl of TE buffer and incubated at 37°C for 10 minutes with occasional shaking. The fines were spun down in a microfuge for 30 seconds and the supernatant, containing the DNA, collected. The fines were then resuspended in 20µl of TE and incubated as before. Further centrifugation and collection of the supernatant gave a total volume of 50µl of DNA.

Preparation of the silica fines:

250ml of silica 325 mesh powder was resuspended in distilled water to give a total volume of 500ml. The suspension was stirred for 1 hour and left to settle for a further hour. The suspension was then centrifuged at 5000g in a Beckman J2-HS centrifuge using a JA-14 rotor. The pellet was resuspended in 150ml of distilled water plus 150ml nitric acid. The suspension was then heated to 98°C and allowed to cool to room temperature. The silica fines were then repeatedly washed with sterile distilled water until the pH was greater than 5.5. Silica fines were stored at 4°C as a 50% slurry in sterile distilled water.

A QIAGEN QIAquick Gel Extraction Kit was also sometimes used, according to the manufacturers instructions, for isolating DNA from agarose gels.

#### **2.5.5 Filling in 3'-Recessed Termini.**

The DNA fragment (maximum of 500ng) was resuspended in 50µl of TE following isolation from an agarose gel. 7µl of a solution containing all 4 dNTPs

(each at 1mM) was added to the DNA. 7µl of Klenow buffer (10x) was added (10x buffer is 0.5M Tris.HCl pH 7.6, 0.1M MgCl<sub>2</sub>) and the reaction buffer made upto 70µl with sterile distilled water plus 1µl (1 unit) of Klenow fragment. The reaction mixture was left at room temperature for 30 minutes and then the Klenow fragment either inactivated by incubating at 70°C for 5 minutes, or removed using a QIAGEN PCR Purification Kit, according to the manufacturers instructions.

**2.5.6 Ligation of DNA.**

T4 DNA ligase was used to ligate DNA fragments with compatible cohesive or blunt termini. The fragments of insert and vector DNA were usually mixed at a ratio of 3:1 (insert:vector) with a maximum of 300ng DNA. 0.1 volume of 10x ligase buffer (0.66M Tris.HCl pH7.5, 50mM MgCl<sub>2</sub>, 50mM DTT, 10mM ATP) was added and 1 unit of DNA ligase. For cohesive termini the tubes were incubated overnight at 4°C. For blunt-ended termini the reactions were incubated at 16°C overnight. The ligation mix was then used immediately to transform competent *E.coli* cells.

**2.5.7 Polymerase Chain Reaction (PCR).**

All PCR reactions were performed in a Perkin-Elmer Thermal Cycler. Reactions were usually carried out in a total volume of 25µl, in thin-walled, 0.5ml, PCR tubes. Per reaction tube the following components were added:

10x reaction buffer	2.5µl
dNTPs	2.0µl (each at 1mM)
primer 1	1.0µl (at a concentration of 10pmol/µl)
primer 2	1.0µl (as above)
Taq polymerase	0.25µl (1 unit)
template DNA	1.0µl (see below)
25mM Mg <sup>2+</sup>	1-4µl (concentration used varied)
dH <sub>2</sub> O	appropriate to make final volume 25µl

The amount of template DNA used varied, however it was usually approximately 100ng for plasmid DNA, and 500-1000ng for genomic DNA.



To avoid pipetting very small volumes (<1µl) the 10x reaction buffer, dNTPs, primers and Taq polymerase were made into a "Master mix". For n tubes, the master mix consisted of appropriate volumes of the different components to make up n+1 reaction tubes. The mixture was then dispensed into each of the reaction tubes.

Primer design: Primers used were generally designed so as to contain an equal number of G+C's and A+T's, usually 10 of each pair. Care was taken to ensure the primers would not self anneal or hybridise to each other, as this would reduce the efficiency of the PCR reactions.

PCR Protocol: The temperatures and times used in PCR reactions followed the same basic pattern:

94°C	5 minutes	
94°C	1 minute )	- denature
55°C*1	1 minute ) x30	- anneal
72°C	1*2 minute)	- extend
72°C	5 minutes	

- \*1 The annealing temperature used varied as it is set by the composition of the primers; The approximate T<sub>m</sub> for each primer can be worked out using the following equation:

$$T_m = \{[(\text{no. of G} + \text{C}) \times 4] + [(\text{no. of A} + \text{T}) \times 2]\} - 5^\circ \text{C}$$

If the annealing temperatures of the two primers differed, the lower temperature was used in the reactions.

- \*2 The extension time is set by the length of the largest fragment expected from the reactions, e.g. if expect a 3kb fragment the extension time would be 3 minutes.

### 2.5.7.1 Hot Start PCR.

Occasionally "hot start" PCR was used, this reduces the amount of non-specific binding of the primers by compartmentalising the contents of the reaction tube until the denaturation temperature is reached. Half the total volume of the reaction (consisting of the 10x buffer, dNTPs,  $Mg^{2+}$ , primer 1 and  $dH_2O$ ) was put in a tube, a wax bead was then added, using sterile forceps, and the tube heated to  $70^{\circ}C$  for 5 minutes to melt the wax. After heating, the tube was held on ice for 2 minutes to solidify the wax and seal off the bottom half of the tube. The second half of the reaction volume (consisting of the Taq polymerase, template DNA, primer 2 and  $dH_2O$ ) was then added, on top of the wax seal, and the tubes placed in the thermal cycler.

## 2.6 Transformation of *E. coli*.

### 2.6.1 Preparation of Competent Cells.

5ml of LM broth was inoculated with DH5 $\alpha$  and grown overnight at  $37^{\circ}C$ , it was then subcultured 1:100 into fresh LM broth and the cells grown to an OD<sub>550</sub> of 0.3-0.35. The culture was chilled for 5 minutes on ice before being spun down in pre-chilled centrifuge tubes at 4000g,  $4^{\circ}C$ , for 7 minutes. The supernatants were poured off and the cell pellets resuspended in  $\frac{2}{5}$  of the original culture volume with solution A (solution A = 30mM potassium acetate, 100mM rubidium chloride, 10mM calcium chloride, 50mM manganese chloride and 15% glycerol. The solution was adjusted to pH 5.8 with 0.2M acetic acid and filter sterilised.) The tubes were held on ice for 5 minutes and spun down as before. The supernatants were poured off and the pellets resuspended in  $\frac{1}{25}$  of the original culture volume with solution B (solution B = 10mM MOPS, 75mM  $CaCl_2$ , 10mM  $RbCl_2$  and 15% glycerol. The solution was adjusted to pH 6.5 with KOH and filter sterilised.) The tubes were left on ice for 15

minutes, then 150µl aliquots were added to pre-chilled eppendorfs, frozen in liquid nitrogen and stored at -80°C.

### **2.6.2 Transformation Procedure.**

The cells were thawed by hand and placed on ice for 10 minutes. The DNA to be transformed was added, and the tube held on ice for 30 minutes. The cells were heat shocked at 42°C for 90 seconds, held on ice for 3 minutes and then 800µl of pre-warmed (to 37°C) LM broth were added. The tube was incubated for 1 hour at 37°C, with continuous shaking. Finally appropriate aliquots, usually 1/10th and 9/10ths of the tube, were spread onto selective agar plates and incubated overnight at 37°C.

## **2.7 DNA Hybridisation Procedures.**

### **2.7.1 Radio-Labelling of DNA Fragments.**

DNA fragments were labelled with [ $\alpha$ -<sup>32</sup>P] dCTP by the random primer labelling method using an Amersham Multiprime kit. 30-50ng of the DNA to be labelled, in a total volume of 28µl, were boiled for 5 minutes and then held on ice for 2 minutes. 10µl of labelling buffer, 5µl of random hexanucleotide primers, 5µl <sup>32</sup>P-dCTP (equivalent to 50µCi) were added, followed by 2µl of Klenow enzyme. The labelling reaction was left to proceed either at room temperature overnight, or for 2-3 hours at 37°C, the probe was then purified through a 10cm Sephadex G-50 column using STE as the elutant. The labelled DNA was collected and stored at -20°C until required. Immediately before use the probe DNA was boiled for 5 minutes.

### **2.7.2 Southern Blotting.**

DNA was transferred to Hybond-N (Amersham) nylon membranes, according to the manufacturer's instructions. Firstly the agarose gel containing the DNA samples was photographed with a ruler down its side. Gels known to have DNA fragments greater than 10kb in size were soaked in 0.25M HCl for 15 minutes, to partially depurinate the DNA, and rinsed twice with distilled water. For blotting the gel was soaked in denaturation buffer (1.5M NaCl, 0.5M NaOH) with occasional

shaking for 30 minutes. The gel was then rinsed twice with distilled water and soaked in neutralisation buffer (1.5M NaCl, 0.5M Tris.HCl pH7.2, 0.001M EDTA). After rinsing the gel twice with distilled water, the blot was set up.

For single sided (one-way) blots a reservoir of 20x SSC was set up (20x SSC - 3.0M NaCl, 0.3M Na.citrate pH7.0). A platform was placed over this reservoir and a long piece of Whatman 3MM paper (presoaked in 20x SSC) put on the platform with its ends dipping into the reservoir. The gel was placed, wells uppermost, on the 3MM paper, and a piece of Hybond-N nylon membrane cut to the same size as the gel placed on top of it. Any air bubbles were carefully removed before 3 sheets of Whatman 3MM paper, cut to the gel size and presoaked in 20x SSC, were placed on top. Finally 2 layers of disposable nappies (also cut to gel size) were placed on top, the stack covered with a glass plate and a 1kg weight placed on top.

Double sided (two-way) blots were created by sandwiching the agarose gel between two equivalent stacks of a glassplate, 2 layers of disposable nappies, 3 pieces of Whatman 3MM paper (presoaked in 20x SSC) and a piece of Hybond-N nylon membrane. A 1kg weight was placed on the top glass plate. Liquid retained in the gel transferred the DNA onto both membranes.

Both types of blot were left for at least 16 hours to allow DNA transfer. After which time the apparatus was dismantled, and the positions of the wells marked on the nylon. The nylon filter was washed carefully in 2xSSC, then allowed to air dry for up to 1 hour before being wrapped in clingfilm. It was then exposed to UV light for two minutes each side to fix the DNA to the membrane.

### **2.7.3 Hybridisation of Radio-Labelled Probes to Blots.**

Hybridisation reactions were carried out using Techne Hybridisation tubes in a Techne Hybridiser HB-1 oven. The nylon filter was put inside the hybridisation tube and 200µl of pre-hybridisation solution (5x SSC, 5x Denhardt's solution [50x Denhardt's solution is 1% ficoll, 1% polyvinylpyrrolidone, 1% BSA fraction V], 0.5% SDS, 0.1% pyrophosphate and 100mg.ml<sup>-1</sup> of denatured salmon sperm DNA) added per cm<sup>2</sup> of filter. Any air bubbles were carefully removed and the tube incubated at 65°C, whilst being rotated in the oven. After two hours the labelled probe was denatured, by boiling for 5 minutes, and 100-200 µl added to the tube contents. The tube was replaced in the oven and incubated at 65°C for at least 12

hours. After incubation the pre-hybridisation solution (containing the probe) was poured off and the blots washed.

#### **2.7.4 Washing of Probed Blots.**

The nylon filter was washed within the hybridisation tube and was never allowed to dry out. The filter was washed twice in 2x SSC, 0.1% SDS for 10 minutes at room temperature, followed by one wash in 0.1x SSC, 0.1% SDS for 15 minutes at 65°C. After each washing solution was removed, the filter was checked with a Geiger counter and the washing continued until sufficient (apparent) non-specific radio-labelled probe was removed (until a reading of <20cpm was obtained). Finally the filters were wrapped in clingfilm.

#### **2.7.5 Detection of Hybridising Probes.**

The wrapped filter was taped onto a larger piece of Whatman 3MM paper in a lead cassette. Radioactive bands could be detected on the filter by exposing Fuji RX-100 X-ray film to the filter. The film sheets were pre-flashed once to sensitise them, and exposure carried out at -80°C, for varying amounts of time. Exposed films were developed with Ilford Phenisol developer for up to 4 minutes and fixed with Kodak Unifix fixer for 2 minutes. The positions of the wells on the filter were marked on the film and the size of any hybridising fragments calculated using the original gel photograph (with ruler included).

#### **2.7.6 Stripping of Probed Blots.**

Blots were stripped by pouring a boiling solution of 0.1% SDS over the membrane and allowing the solution to cool to room temperature. Stripping was usually allowed to proceed for a period of at least two hours, after which time the membrane was wrapped in clingfilm and autoradiography carried out for the normal exposure time to check that the probe had been completely removed. The membrane was then pre-hybridised and hybridised with a new probe.

## **2.8 DNA Sequencing.**

DNA sequencing was carried out with an Applied Biosystems 373A DNA Stretch Sequencer and a 377 DNA sequencer, using double stranded DNA templates and dye terminator chemistries. Usually, the universal M13 forward and reverse primers were used, but occasionally custom-synthesised oligonucleotides were also used.

Nucleotide sequence searches were carried out against the GenBank and EMBL databases using the Fasta programme at SEQNET. Putative protein sequences were searched against the Owl database, using the programmes Sooty and Sweep, and BLASTP.

## **2.9 Conjugation of Plasmids into *Agrobacterium*.**

Triparental matings based upon the method of Ditta *et al.*(62) were used to mobilise plasmids into *Agrobacterium* with pRK2013 as a helper plasmid. Cultures of the recipient *Agrobacterium* strain, and the *E.coli* plasmid donor and helper strains were grown to mid log phase. 100µl of each culture were pipetted onto a 0.22µm nitrocellulose filter disc on the surface of an LM-agar plate. This was incubated at 28°C overnight, and the disc then transferred to a universal bottle containing 10ml of 10mM MgSO<sub>4</sub>. This was vortexed vigorously to wash the bacteria off the disc. Dilutions were made from the resulting cell suspension and plated onto agar plates, with appropriate antibiotic selection.

## **2.10 Mutagenesis.**

### **2.10.1 Gene Replacement Mutagenesis.**

Gene replacement mutagenesis was carried out according to the protocol of J. Quandt and M. F. Hynes (162). The technique involves replacement of a functional gene with a copy of the same gene interrupted by the insertion of an antibiotic resistance cassette, thereby creating a non-functional copy of the gene in question.

Two vectors were used in this work, pJQ200mp18 and pJQ200uc1, which only differ in the restriction enzymes available in their respective multiple cloning sites. The pJQ200 vectors have gentamycin resistance and a  $\beta$ -galactosidase gene for selection procedures. In addition they have a functional *mob* site which allows conjugation into *A. tumefaciens* strains. They also possess the *sacB* gene from *B. subtilis*, which when activated by sucrose, produces levansucrase, an enzyme whose expression causes the production of toxic compounds that are lethal in Gram negative bacteria. Also crucial to their functioning in this work, is the fact that the pJQ200 vectors are unable to replicate outside of the enterobacteria, they therefore act as suicide vectors in *A. tumefaciens*.

Genes to be mutated were ligated into a pJQ200 vector and correct insertion confirmed by inactivation of the  $\beta$ -galactosidase gene. The neomycin resistance cassette from pDUB2033 was then ligated into the gene and positive constructs obtained by selection on plates containing both gentamycin and neomycin. The final plasmid construct was then conjugated into *A. tumefaciens* C58C1 by tri-parental mating (see section 2.9).

The pJQ200 constructs are only able to survive in *A. tumefaciens* by integration into the genome via homologous recombination. For correct construction of the mutant strain, a rare second recombination event must occur between the mutated gene and the wild type copy, thus excising the wild type copy of the gene and the plasmid vector from the cells.

After the first recombination and subsequent integration step the entire mutant construct was incorporated into the genome, giving the bacteria both gentamycin and neomycin resistance. Since both intermediate strains and the final mutants possess resistance to neomycin it was not possible to select for correct construction of the mutant by simply growing on plates with neomycin selection alone. The presence of the *sacB* gene on the pJQ200 vector however acts as an additional selectable marker, allowing direct selection of only those strains that have undergone the second recombination event, and contain only the mutated copy of the gene. Selection on sucrose induces the *sacB* gene, thereby killing those strains that still contain the vector sequence.

Tri-parental matings were therefore plated on LM-agar plates containing rifampicin, neomycin and 5% sucrose.

To confirm that gene replacement had occurred, genomic digests and subsequently Southern blots were made of the putative mutant strains and the wild type strain C58C1. The blots were probed with a radioactively labelled copy of both the wild type gene, and the neomycin cassette. Due to the insertion of the neomycin cassette, when digested with the same enzyme, different sized hybridising bands were expected in the mutant strains, as compared to the wild type strain, when probed with the gene in question. The neomycin probe should only hybridise with known sized fragments in the mutants. The differences in band patterns between the wild type strain and the mutant strain therefore confirms gene replacement has occurred.

## **2.11 Bacterial Growth Measurement.**

The optical density of bacteria was measured at 600nm on a Beckman DU7500 spectrophotometer to produce growth curves of the changes in optical density over time.

## **2.12 Microscopy.**

### **2.12.1 Light Microscopy.**

A loopful of bacteria, taken from a fresh swarm plate, was resuspended in 40µl of chemotaxis media (1ml 0.1M EDTA and 10ml 1M  $\text{KH}^2\text{PO}_4$  (pH7.0) made to 1 litre with distilled water.) and observed under phase contrast optics (using a Nikon Optiphot microscope).

### **2.12.2 Electron Microscopy.**

A loopful of bacteria, taken from a fresh swarm plate, was washed onto a Formvar-coated grid with 100µl of 10mM  $\text{MgSO}_4$ . The bacteria was allowed to settle on the grid for 5 minutes before the excess solution was blotted off. Grids were not allowed to dry out completely at this stage. A drop of saturated aqueous uranyl acetate (pH 4.8) was placed on the grid for 30 seconds, the excess stain was then partially blotted off and the grid allowed to air dry. The grids were viewed and photographed using a Philips EM400.





### **3. Gene Identification**

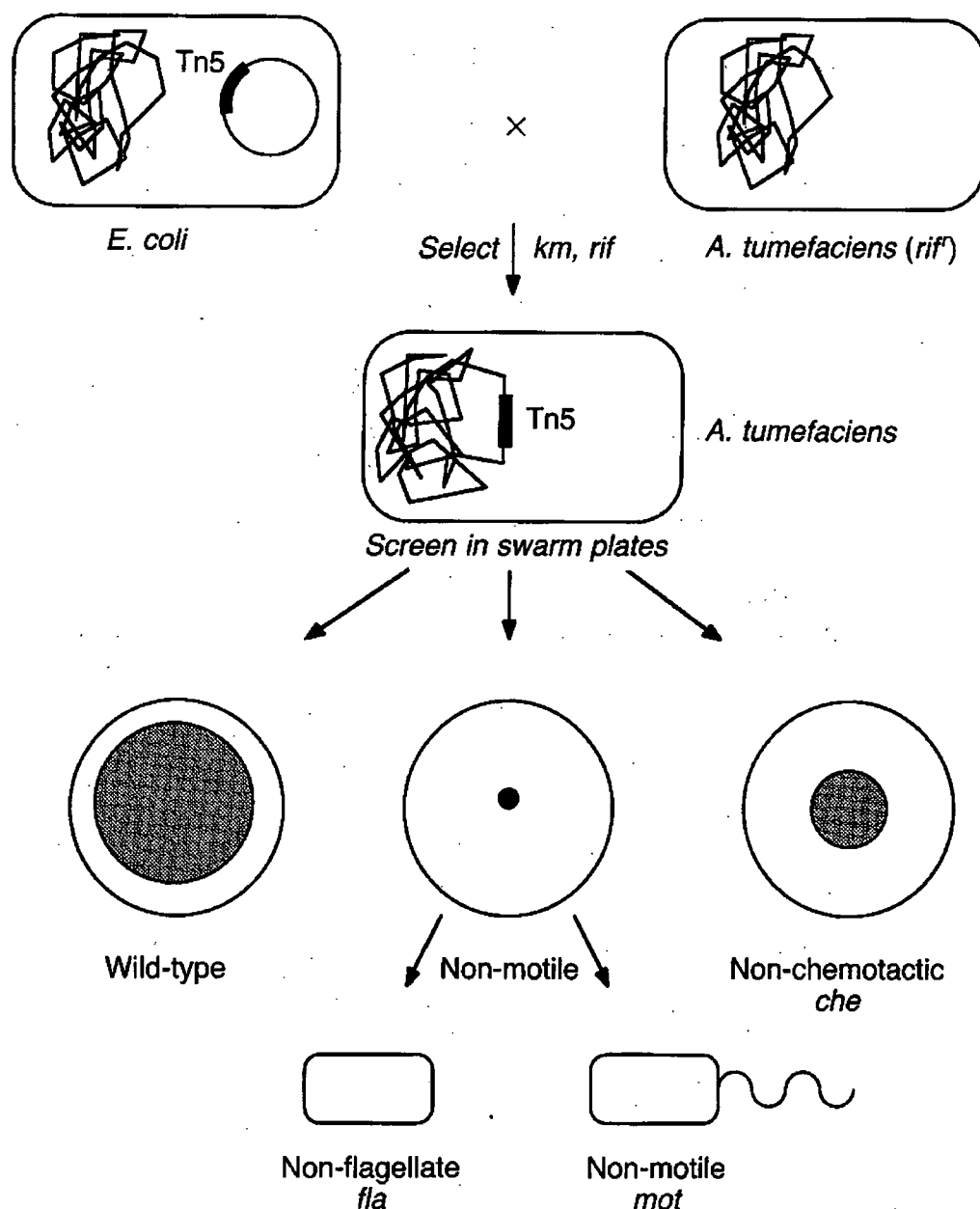
### 3.1 Background of this project:

In order to study motility and chemotaxis genes in *Agrobacterium tumefaciens* in more detail, a number of mutants were generated using a Tn5 suicide vector, followed by selection for non-motile and non-chemotactic mutants. Tn5 is a 5.8 kb transposon carrying a kanamycin resistance gene and has specific properties that facilitate genetic analysis of resulting mutants (185). Tn5 has transposase activity allowing it to insert itself randomly, with a high frequency, into the chromosome or plasmid of a wide range of Gram-negative bacteria, producing a mutation at the insertion site. It exhibits a low probability of genome rearrangement during insertion and a high stability when inserted. The DNA fragment, containing Tn5 can be easily isolated using standard methods.

A plasmid with a high transfer frequency was constructed to deliver the insertion mutagen Tn5 (117). The *E. coli* strain carrying the Tn5 "suicide vector" was mated directly to the rifampicin resistant *A. tumefaciens* strain C58C<sup>1</sup>. Transconjugants were screened for motility defects in L-agar swarm plates containing Rif and Km antibiotics. Putative motility and chemotaxis mutants were isolated by selecting strains that produced smaller swarm morphologies compared to the wild type (figure 3.1.1) (180).

Twenty independent behavioural mutants, showing impaired swarming ability, were isolated (180). They were further characterised using flagella staining, electron microscopy and motility studies with a phase contrast microscope. Preliminary phenotype analysis showed that 14 of the non-swarming mutants were non-motile. One of the mutants however proved to be motile but tumbled continuously, and was thus identified as a possible non-chemotactic mutant and assigned as *che-2*. Although *che-* mutants are fully motile, they are thought to form smaller swarms than wild type due to a defect in the intracellular signaling pathway.

Electron microscopic observation divided the 14 non-motile mutants into two subclasses: those possessing normal flagella, which are presumably non-functional in some flagellar components (*mot-* mutants) and those lacking flagella, which were designated *fla-*. The other mutants, *sww* and *tpc* have flagellar structures different from the wild-type, however they show poor motility and small swarms (180).



**Figure 3.1.1 Tn5 mutagenesis.** (see text for details). Adapted from(180).

Further experiments were carried out to clone the DNA fragments surrounding the Tn5 insertion sites. Chromosomal DNA from each mutant was isolated, digested with *Eco*RI, then probed with a radioactively labelled Tn5 fragment. The bands containing the Tn5 insertion site plus flanking sequences from each *A. tumefaciens* behavioural mutant were cloned into the plasmids pUC18/19. The new plasmids were designated pDUB1801 to 19. The mutant flanking sequences were isolated by cleavage

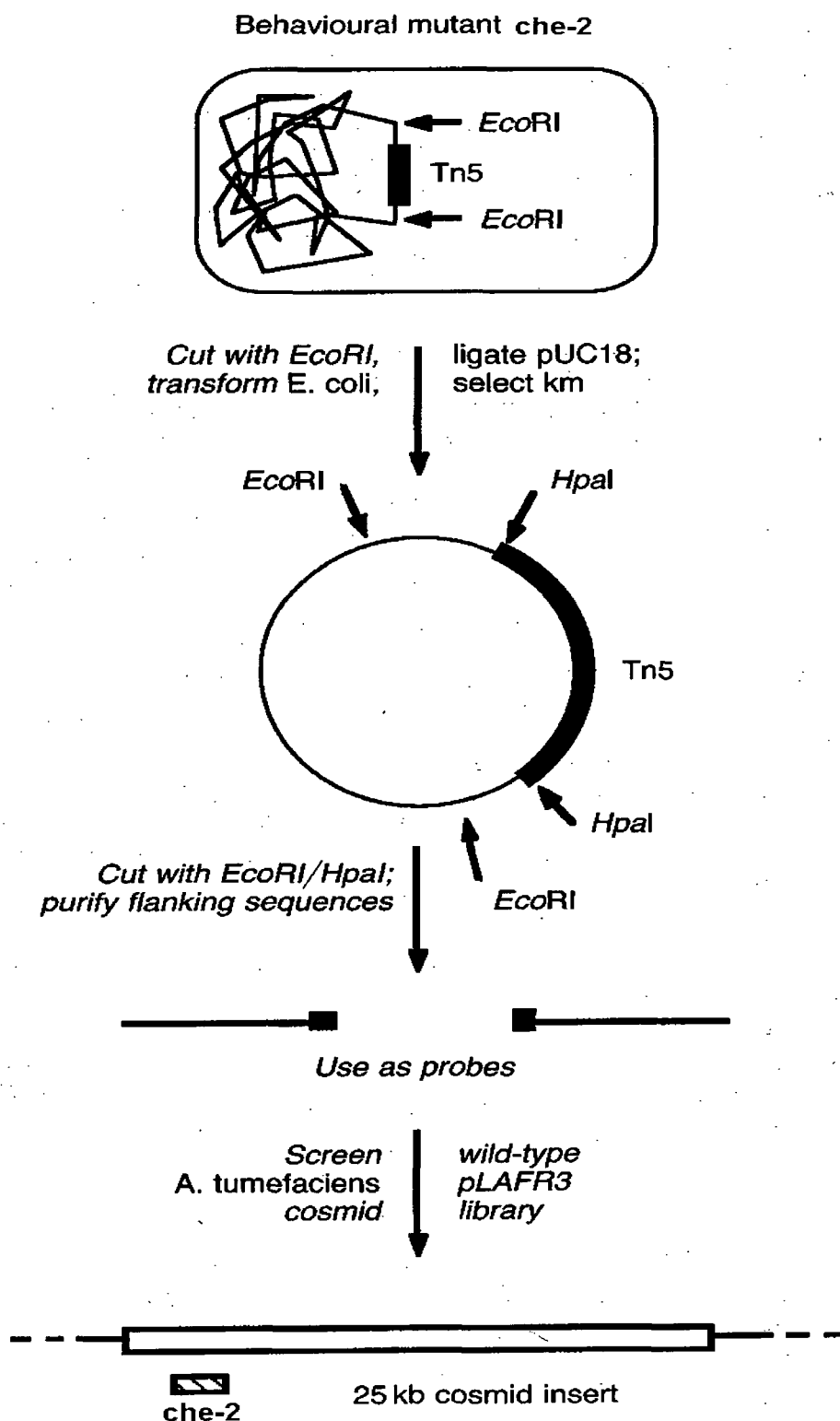


Figure 3.1.2 Isolation of behavioural mutants. *che-2* is shown as an example (see text for details). Adapted from (180).

with *EcoRI* and *HpaI*. The latter enzyme has two sites, 185 bp from each end of Tn5, while *EcoRI* has no site in Tn5.

To isolate DNA that codes for chemotaxis and motility genes in wild-type *A. tumefaciens*, a representative genomic library of C58C<sup>1</sup> genomic DNA was constructed (117). *A. tumefaciens* C58C<sup>1</sup> DNA was partially digested with *Bam*HI and 20-25 kb fragments ligated into the cosmid vector pLAFR3. The ligation mixture was transformed into *E. coli* and 1200 separate recombinant clones were selected and maintained separately. The motility flanking sequences were then radiolabelled and used as probes to screen the genomic library cosmids (see figure 3.1.2). The identified cosmid clones were then used in complementation tests on the mutants, and any others found to hybridise to the same cosmid. This method showed that the mutated loci mapped to two separate cosmids carrying a total of 12 loci. Cosmid clone pDUB1900 was found to either complement or hybridise to seven of mutants: *mot-1*, *mot-4*, *mot-9*, *mot-12*, *fla-3*, *fla-11* and *fla-15*. Cosmid clone pDUB1905 was found to complement three mutants, *che-2*, *mot-6*, and *fla-8*. The two latter mutants share the same position in the cosmid pDUB1905. E. L. Wright has recently constructed a cosmid, pDUB1911, using heterologous probing with a DNA fragment from *S. meliloti*. pDUB1911 contains a *che* operon which consists of a number of chemotaxis and motility genes (228).

### 3.2 Sequencing plasmid pH1.

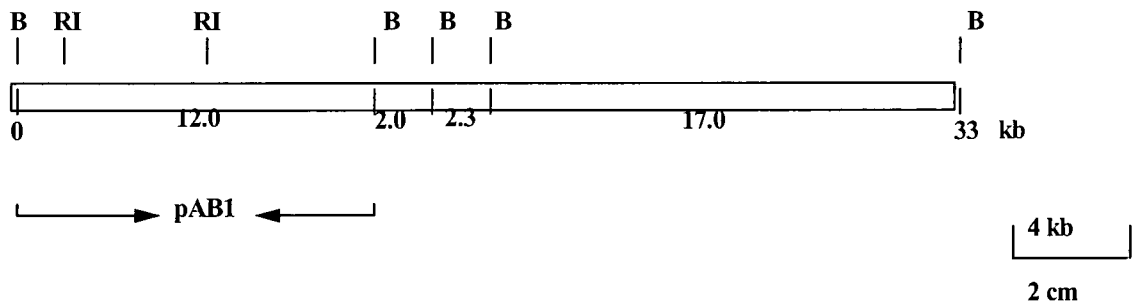
It has previously been shown that the behavioural genes in the cosmid pDUB1905 are located on a 12 kb *Bam*HI fragment. This fragment was isolated to produce the plasmid pAB1. The initial work was to sequence a 5 kb *EcoRI* fragment from pAB1. This region was known to contain DNA complementing three phenotypically distinct motility and chemotactic mutants, *mot-6*, *fla-8* and *che-2* (see figure 3.2.1a, and 3.2.2)(180).

Generally, DNA fragments to be sequenced were subcloned into pBluescript SK<sup>+</sup> and the universal M13 forward and reverse primers were used as described in section 2.8. SK<sup>+</sup> was chosen because of its wide range of unique restriction enzymes that make analysis of double digestion easier, and the *lacZ* region, which is used for blue/white selection of recombinant plasmids.

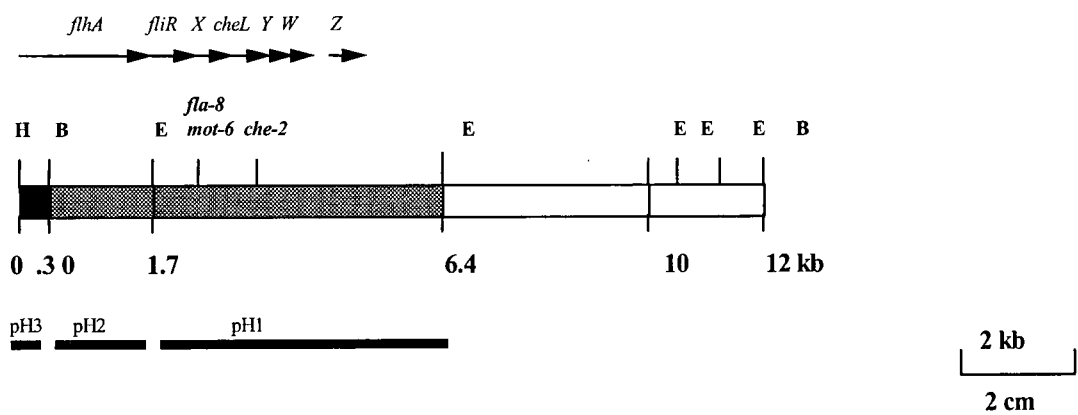
### Figure3.2.1 Cosmid pDUB1905 and major plasmids derived from the cosmid.

- a) Cosmid pDUB1905. The arrows underneath the cosmid showing plasmid pABI.
  - b) Plasmid pABI, arrows above the plasmid indicating the position of the genes and ORFs found in this work. X, Y, W and Z representing *orfX*, *Y*, *W*, and *Z*. Arrows underneath pABI are showing three other plasmids generated from pABI.
  - c) Plasmid pH1 and three major plasmids, which generated in this work.
- Positions corresponding to the mutations and restriction enzyme also have been shown.
- B= *Bam*HI   E or RI=*Eco*RI   R5=*Eco*RV   H=*Hind*III

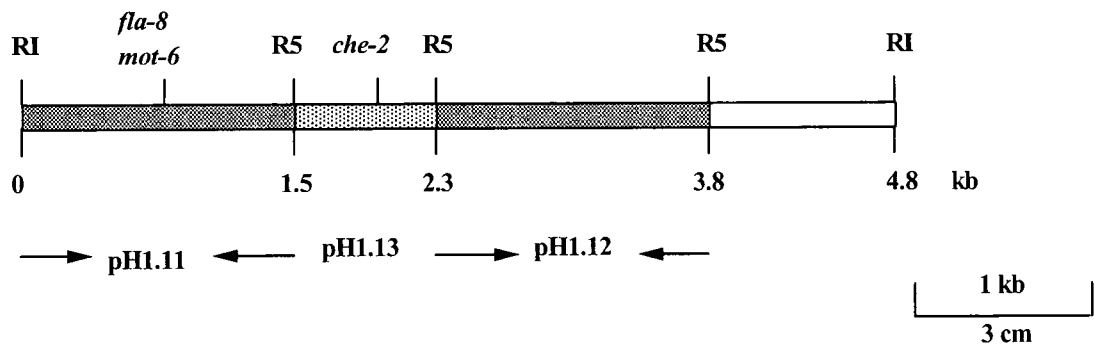
a.



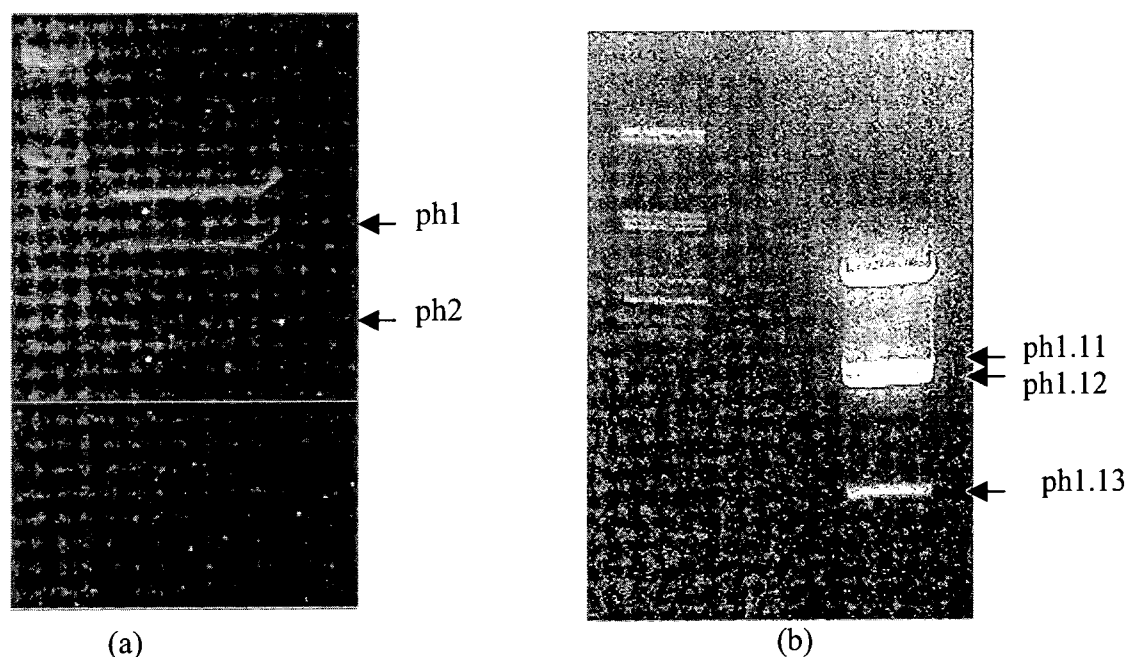
b.



c.







**Figure 3.2.2 Double digestion of pAB1 (a) with *Bam*HI-*Eco*RI , and (b) pH1 with *Eco*RI-*Eco*RV.**

The fragments isolated and used for subcloning are shown by arrows.

The 5 kb *Eco*RI fragment was subcloned into pBluescript SK<sup>+</sup>, producing the plasmid pH1. Sequencing of this plasmid was carried out by further subcloning of fragments isolated from the plasmid pH1 into pBluescript SK<sup>+</sup>. When there were no convenient sites for restriction digesting, primers were designed for sequencing. The plasmid pH1 has been fully sequenced in both directions (see figure3.2.1c).

Raw sequence data was initially manipulated and aligned using the Macintosh computer programs DNA Strider<sup>TM</sup> 1.2 and Sequencher. Restriction enzyme sites were found by computer analysis, and confirmed by experimental digests of the DNA and comparing the size of the fragments produced with those predicted. Possible open reading frames (*orfs*) of the pH1 sequence were identified using DNA Strider and the Seqnet program TESTCODE (this program calculates the probability of the sequence to contain coding regions.).

The whole sequence was compared against nucleic acid databases to identify homology for chemotaxis or motility genes. The University of Wisconsin genetics computer group (UWGCG) sequence analysis software package mounted on the SEQNET VAX 3600 at SERC, Daresbury, UK, was used for DNA homology searches. The sequence was compared against the GenBank and EMBL databases using the FASTA program. Alternatively the *orfs* were first translated into their protein sequence and compared against the Owl database using the SEQNET program SWEEP and BLASTP.

Preliminary TESTCODE analysis of pH1 showed that the sequence contained several coding regions. A number of possible open reading frames, in both the sense and antisense directions were also identified. Subsequent DNA and protein searches using the open reading frames of coding regions showed significant sequence homology to known motility and chemotaxis genes of many bacteria. The 5' end of the sequenced region was a partial *orf* that when translated, showed homology with FlhA from many bacteria. Other ORFs displayed substantial sequence similarity to FliR, and CheL in the databases (see figure 3.4.2).

### 3.3 Sequencing pH2 plasmid.

The open reading frame which showed homology to FlhA contained only 120 amino acids, whilst this protein in other bacteria consists of about 680 to 707 amino acids. This result suggests that the open reading frame extend further in an adjacent fragment. To find the beginning of this open reading frame, a 1.7 kb *Bam*HI-*Eco*RI fragment, upstream of the pH1 fragment from plasmid pAB1, was subcloned into pBluescript SK<sup>+</sup> (see figure 3.2.2b). This created a plasmid named pH2. The plasmid was fully sequenced in both direction and aligned to the pH1 sequences (see figure 3.2.1b). TESTCODE analysis of the new alignment showed that the whole sequence of pH2 with about 350 bp of 5' end of pH1 contained a large coding region. This result was further confirmed when open reading frame maps of the alignment, showed a large *orf* in this region. Homology analysis also showed a potential similarity between this ORF and FlhA of many bacteria (see table 3.5.1), however the beginning of the *Agrobacterium flhA* sequence was still missing. Next step of project, therefore was construction of missing part of *flhA*.

### 3.4 Construction and sequencing of plasmid pH3.

The 1.7 kb *Bam*HI-*Eco*RI fragment pH2 (as can be seen in figure 3.2.1) includes the 5' end of the cosmid pDUB1905. This, together with the results obtained from homology searches, suggested that the beginning of the open reading frame showing homology to *flhA* gene was in the DNA fragment adjacent to the cosmid pDUB1905. Previous work conducted in this laboratory showed that cosmid pDUB1905 does not overlap with cosmid pDUB1900, which contains several flagellar genes. Therefore, reverse PCR was used to construct the beginning of *flhA* (see figure 3.4.1). A Southern blot experiment was carried out to find the appropriate restriction enzyme for digesting the *Agrobacterium* chromosomal DNA.

*Agrobacterium* genomic DNA was digested with different restriction enzymes include *Hind*III, *Eco*RI, *Hpa*I, *Hinc*II, *Bcl*II and *Sac*I. The digests were probed with a 0.8 kb pH2/*Hind*III fragment from pDUB1905. The result showed that *Hind*III is the most suitable site to use for reverse PCR.

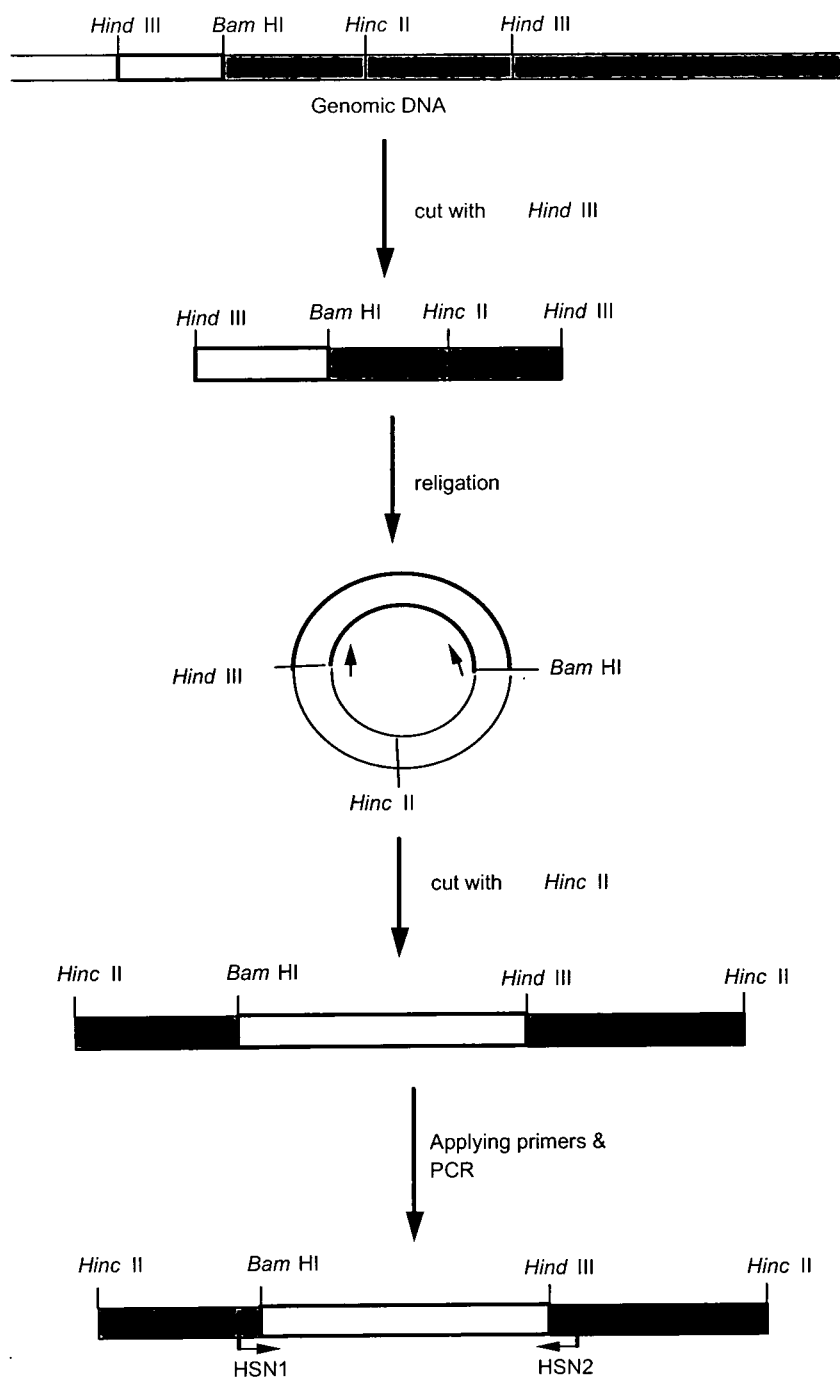
Reverse PCR is used when the sequence of only one end of a fragment is known. In this method specific primers are designed the opposite direction. Two specific primers, which were used in the PCR, were designed from pH2. They were:

HSNI: 5'-TTC AGC GAC AGA CGG ATC AT-3'  
57 76

HSNII: 5'-GCG GCA TCA TCA TCG GTT AT-3'  
476 495

Using these oligonucleotides in the PCR reaction allowed synthesis of chromosomal DNA directly upstream of the 5' end of the cosmid pDUB1905.

The diagram of reverse PCR is shown in figure 3.4.1. *Agrobacterium* genomic DNA was digested with *Hind*III. The digest was religated and then cut with *Hinc*II to open the circle. Two specific primers (HSNI and HSNII) were applied to the PCR reaction samples containing the digest. The PCR program, described in section 2.5.7, was used. An approximately 0.3 kb fragment was synthesised. The PCR product was subcloned in SK<sup>+</sup> and called pH3. The complete length of the plasmid was sequenced in both directions. The plasmid pH3 sequence overlapped with the 5' end of pDUB1905, meaning that it was indeed a continuation of pDUB1905. Therefore, The pH3 sequence was aligned to the other sequences, creating a new alignment about 5.8 kb in length (see figure 3.2.1b). The complete sequence of the sense strand of 5.8 kb



**Figure 3.4.1** Diagram outlining the main steps involved in reverse PCR to synthesise the beginning of *flhA*.

The box with thick borders shows the fragment to be synthesised. Arrows within the circle and underneath the main figure indicate the primers and the direction that they act to synthesise the beginning of *flhA*.

Dark shaded regions= pDUB1905.

**Table 3.4.1 Some characteristic of potential genes and *orf* found in this work**

Name	Nucleotide Length (bp)	Protein length (aa)	Molecular weight (kDa)	Isoelectric point (Ip)	G+C% content
<i>FlhA</i>	2109	702	76.865	6.679	59.03
<i>FliR</i>	756	251	26.809	7.143	57.54
<i>OrfX</i>	420	139	16.053	5.931	54.76
<i>CheL</i>	432	143	15.565	4.994	54.86
<i>OrfY</i>	396	131	14.722	4.910	54.55
<i>OrfW</i>	453	151	16.497	4.920	57.40
<i>orfZ</i>	675	224	24.588	8.712	60.89

**Figure 3.4.2 The complete sequence of the sense strand of 5.8 kb region of the cosmid pDUB1905.**

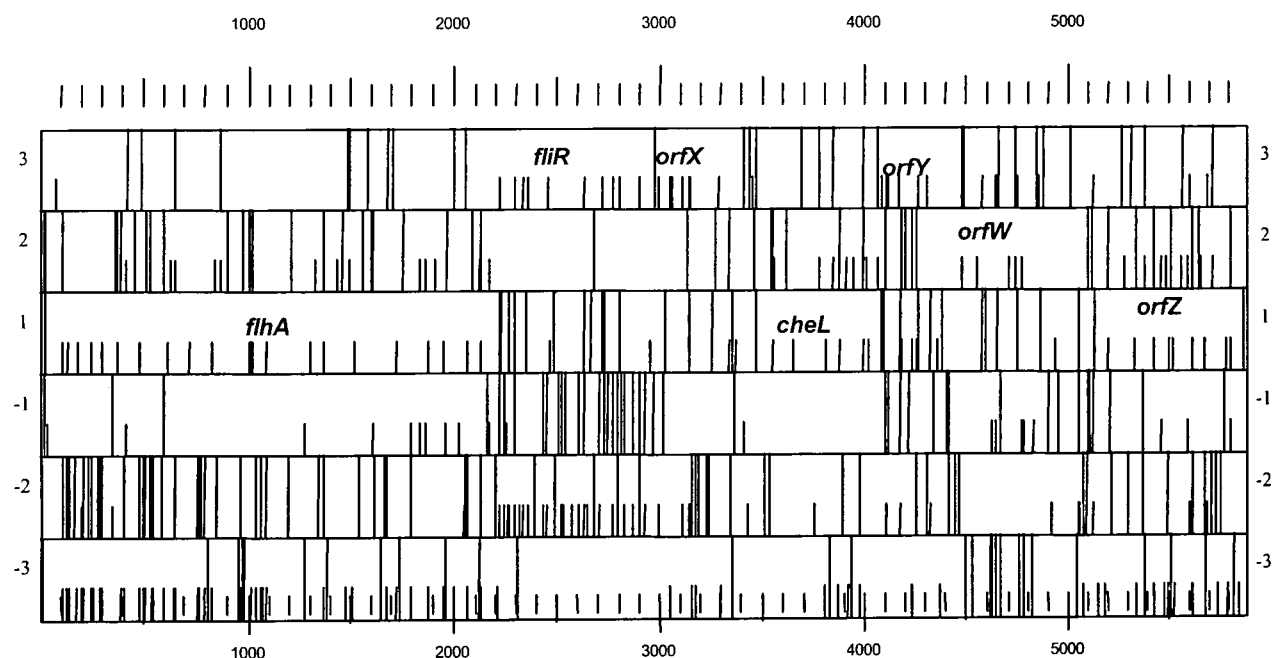
	5	15	25	35	45	
1	AAGCTTCAGC	GGCTAGGGAT	AGGCTGAAGG	CATCGGCCCG	AATTGCGTTC	50
51	GCTTTTCGAA	AACCGATGCC	CGGAACCGTG	CGACGGCTTT	TCCGTCGCCC	100
101	TTCTCATGAA	GAGACAGGAC	GAATTCATGG	CGCAGCCACC	AGTCATCTCC	150
151	TTGCCCAAGG	TCAGTCCGAG	CATGCGGGAT	ATCGGTTTCG	CATTGGGCAT	200
201	CATTTTCGATC	CTTTGCGTTC	TGTTTCCGCC	TATTCCGGTC	ATGCTGGTGG	250
251	ACATCGGTCT	GGCGTTCTCG	ATTACGCTTT	CGGTCCTCAT	CCTCATGGTG	300
301	GCGCTCTGGA	TCCAGCGCCC	GCTGGACTTC	TCGTCATTCC	CGACCGTGCT	350
351	GCTGATTGCC	ACAATGATCC	GTCTGTCGCT	GAACATTGCG	ACGACGCGTG	400
401	TCATCCTTTC	GCATGGTAAC	GAGGGGCCGA	CGGCGGCAGG	TGGCGTGATC	450
451	GCGGGTTTCT	CCAGCCTCGT	CATGTCCGGT	GACTTCGTCA	TCGGTCTCAT	500
501	CGTCTTCCTG	ATCCTGATCA	CGGTCAACTT	CATCGTCATC	ACCAAGGGTG	550
551	CGACGCGTAT	CGCCGAAGTC	GGCGCGCGCT	TCACCCTAGA	CGCCATCCCC	600
601	GGCAAGCAGA	TGTCGATCGA	TGCGGATCTG	TCCGCCGGCA	TCATCGATGA	650
651	AAAGGAAGCG	CAGCGTCGCC	GCCGGGAACT	GGAAGAAGAA	AGCTCTTTCT	700
701	TCGGTTCGAT	GGACGGTGCC	TCGAAATTCG	TCCGCGGCGA	CGCGATCGCC	750
751	GGCCTCATCA	TCACCGCGAT	CAACGTTTTC	GGCGGCATCA	TCATCGGTTA	800
801	TTTCCGCCAC	GGCATGCCGA	TCGGCGAAGC	GGCCGATGTT	TTCGTCAAGC	850
851	TTTCGGTCGG	TGATGGTATC	GTCTCGCAGA	TCCCGGCCCT	GATCGTATCG	900
901	CTGGCGGCCG	GCCTTCTGGT	GTCGCGCGGT	GGCACAGCCG	GATCGACGGA	950
951	TCAGGCCGTC	ATCAATCAGC	TGAGCGGTTA	TCCCCGCGCT	TTGATGGTCT	1000
1001	CGGCCATGCT	GATGGGTCTG	CTCGCCGTCA	TTCCGGGCCT	TCCCTTCGTG	1050
1051	CCCTTCATTT	TCCTCGGCGG	CATCATGGCC	TTTGGCAGCT	GGTATATTCC	1100
1101	GCGCCAGGCC	GAAGCCGAAA	GCGCGCTTCG	CCGCCAGGAA	GAGGAAAACA	1150
1151	AGGTTCTCCA	GACCACCGAG	GCGGAAAAGG	ATTCCGTCAA	GCAGGTGCTG	1200
1201	AAGACCGCCG	AGATCGAGCT	CGCGCTCGGC	AAGCAGGTTT	CCACCCGCCT	1250
1251	TCTCGGCGCG	CATCAGGAAC	TGGCCTTCCG	CGTCGGCAAG	ATGCGCAAGA	1300
1301	AATTCGCCAC	GCAATATGGC	TTTGTGGTGC	CGGAGATCAA	GGTCTCCGAC	1350
1351	GACATCATGA	TCCCGGAAAA	GGCCTATCAG	ATCCGCGTCC	ACGGCACGAC	1400
1401	GATCGCGTCC	AGCAACCTGC	GCGTCGGCGA	TGTGCTGGTC	GTGACGGGGG	1450
1451	CAGGGCGCAA	GCCGAGCATC	CCAGGTGATG	AAATCCGCGA	ACCCGCTTTC	1500

1501	GGCATGCCGG	CCGTCTCGAT	CCTCGAGACC	TTTACCGAGG	ATCTGAAGCG	1550
1551	CGAAGGCTTC	CACCCGATCG	ATAACGTCTC	TGTCGTGCTG	ACGCATCTGA	1600
1601	GCGAGGTTAT	CCGCAACAAT	CTGCCGCAGC	TTCTGTCCTA	CAAGGACGTC	1650
1651	AAGATCCTCA	TCGACAGGCT	TGATCCCGAA	TACAAGAAGC	TGGCTGACGA	1700
1701	GATTTGTTTCG	TCGCACATGT	CCTATTCCGG	CCTGCAGGCG	GTGCTGAAGT	1750
1751	TGCTGCTGGC	CGAGCGCGTA	TCGATCCGCA	ACCTGCATCT	CATTTTGGAA	1800
1801	GCGGTGGCAG	AACTCGCACC	GCATGTGCGC	AAGACCGAGC	AGATCGTCGA	1850
1851	ACATGTCCGC	GTTCGCATGT	CGCAGCAGCT	CTGCGGCGAC	CTTGCCGACA	1900
1901	ATGGCGTGCT	GCGCGTCCCTG	CGACTCGGCA	ACAAGTGGGA	CATGGTCTTC	1950
1951	CATCAGGCGC	TGAAACGCGA	CCAGAAGGGC	GAAATCGTCG	AATTTGACAT	2000
2001	CGATCCGCGC	CATCTCGAGG	AGTTTTCCGA	GCAGGCATCG	AAAGTTATCC	2050
2051	GTGAATTCAT	GGATCGCGGG	CTGCCCTTTG	TCCTTGTAAC	CTCGCCGGAA	2100
2101	ACGCGGTCCCT	ATGTGCGCAT	GATTATCGAG	CGACTCTTTG	CGACCCTGCC	2150
2151	GGTTCTTTCA	CATGTGGAAC	TGGCCAAGGG	TCTCGAGATC	AAGATTCTGG	2200
2201	GCGCCATTTTC	ATGATAACCG	ACCCGCAGGG	AACAATCATT	GCATTGTTCC	2250
2251	TTGCCATCTG	CCGCATAGGC	GCCTGCTTCA	TGACCATGCC	GGGCTTTTCA	2300
2301	AGTTCGCGCA	TCTCGCCGCA	GATTCGCATG	CTTCTCTGTG	TAGCGGTATC	2350
2351	CATGGCGCTT	CTGCCCCGTCC	TGTGGGACAC	CATCTACCCA	AAGGTTTCCG	2400
2401	GTGCGAGCCA	GGGCGCGGTC	GTTGGCCTCA	TCTTCTCCGA	GGTTCTCATC	2450
2451	GGCGCAATGT	ATGGGCTGAT	CGCCCGTTTC	TATACGCTCG	GTTTCCAGTT	2500
2501	CACCGGTGCG	CTCATCGGCG	CTTCCATCGG	CCTCAGTGCG	CCCGGCGGTG	2550
2551	CCGATCCCAT	CGAAGACGTG	CAGGAAAACC	AGATCGCCAA	TTTCATCACC	2600
2601	TTCGGTGGCC	TGCTGGTGCT	GTTCATGATG	GATTTCCACC	ACATCGTCCT	2650
2651	GAAGGCGCTG	GTCGATTCCCT	ATAGCGCCAC	ACCGGTCGGT	GCGCTCATCA	2700
2701	GCGGACAGAA	GATGCTGATC	ACGCTGACGG	ATACGCTCAG	GGCGTCCTTT	2750
2751	TCGATCATGC	TGCGGCTGGC	GAGCCCCTTT	GTTATTTACG	GCATATGTT	2800
2801	CAACGTGCGG	GTCGGTCTCA	TCAACAAGCT	GGCGCCGCAA	ATCCCGGTCT	2850
2851	TCTTCATATC	GACACCCTTC	GTTCTGGCGG	GTGGTCTTTT	CATGCTTTAT	2900
2901	TTGTCGGTCG	CGGGCCTCAT	CCGGCAATTC	GTGGATGGGT	TCGGCCCCGGT	2950
2951	CTTTATCGGC	TTTTGATCGG	GAGGAAACGC	CATGGCTTCG	GACAAGCGCT	3000
3001	CGGAAAAGCT	CAAGCGTCTG	GTAACGGTCC	AGCGGCATAT	GGAAAAAATG	3050
3051	GCCGAGGTGG	AACTGGCCGA	TACCACTCGC	GTGCGCAGCG	AAGTGGCGCA	3100
3101	ATCCATGGAG	AGCACGTTTG	AGGCCATGAG	TTCCATGGAA	CCGGTGCATC	3150
3151	AGACCTTCTC	CAAGCATTAT	TCGGACCGCT	ACAGCCGCCT	GGTCGTCCAG	3200
3201	GACCGCCAGC	TCGAAGGCGT	TCAGCAGTTT	CAGGAAAACA	AGGTTCTGAA	3250
3251	GGAAAAGACC	AAGGCTGACA	GGCTGGAGGA	CAGGATGCAT	ATTGCCCCGCG	3300

3301	ATCTCGAGGA	TCGCGAGGCG	GGTGATAATG	CAATCTACGA	TTTGTTAGAA	3350
3351	ATTACGAATG	TTTCCCACAC	ACCAGCCTCC	AGCAAGGTTG	GCGATCCATA	3400
3401	GTCTTCCCGT	TGTTGCATCG	CGGGCGTCTG	AAATACACCG	TTCGCGATGC	3450
3451	TTCTTGATTG	AGTGAACGCG	TCGGCTTTCC	TCCGGCACGT	TCGTCAGCTC	3500
3501	GGGAGTTCGG	GACGTGGCGA	TATCGGTTAT	AAGCGATCTG	GTGATGGCAT	3550
3551	GTGGTTCGTC	GCCGGCGGCA	CCCGCAGGAA	GTGCAGGTTG	CCCAGGAAAA	3600
3601	ACTGAAGGCA	AACAAGGCGG	CTTTTGCCGA	ACCAGTCTGG	CCGATGCCGG	3650
3651	CAAGGGCTTT	GGTGCTGCCG	TCGATATGCT	TTGACGGTGC	CTCCTCGAAG	3700
3701	GCCGGCCTTG	GCGATACCAA	TATTCGCTCG	GCACGCACCG	AGATTCCCGA	3750
3751	AACATATCGC	AAATATGAAG	CCTCGGTTCT	CCAGACCTTC	GTTGCCAACA	3800
3801	TGCTGCCGAA	AGACAGCGAG	GAAGTTTACG	GCAAGGGTAA	TGCCGGCGAA	3850
3851	ATCTGGAAGA	GCATGATGGC	GGAACAGTTT	GCCGACACTA	TTTCCAGAAA	3900
3901	TGGCGGCGTC	GGCATTGCGG	AGCAGGCTTA	CAAGGATGCG	CTGCGGAAAG	3950
3951	CCGAAAGCAA	AGGCATTACC	GACGTGTCGA	TGAATGACAA	AGACCACAAT	4000
4001	GCTGCAATTC	GGATGGTGGC	GGAGTTCGAG	CGGCAGGTCC	TCGGCGTTTC	4050
4051	CAATGATAAA	ACGGACGAGG	CTTGAGAATG	AAAAACCTTA	ACGAGGAGAC	4100
4101	AAGCATGGAC	CTTATGTCGA	ACGACCACCG	TATCCAATCC	GTTTTCGGCC	4150
4151	GCCTCGAGAT	GATCATCGAC	AATGAGAACG	ACAATATCGG	TAAGGACCCG	4200
4201	CAGTTCGACT	TCAAGGTCTC	CAATGCCCAT	AAGAGCCGTT	GCCTTTATGA	4250
4251	ACTGACGATG	CTGGTCCGCG	ATACACCGCA	CGAAGATATC	GCCGCCGGTT	4300
4301	ACATGGAGCA	GGTGAAGGGC	ATCAAGTCCA	AGCTCGCCAC	CAATGCCACC	4350
4351	CGCGTCGAGG	CGCATCTGAA	CGCGGTTCGC	GCTGTCGCGG	ATCTGCTCAA	4400
4401	GAACGCTATT	CAGGAAGCCG	ATACGGACGG	CACCTATTCC	CAGGAACAAT	4450
4451	TCCTTTACGG	CGCAGCCTCC	TGATGTTGAA	GCTTCTTCTC	ACCGGCGTCT	4500
4501	GGGTTTGCGC	CGTCACGCTC	GGCGCGGTGT	ATTTCTCCGT	CCAGATGGCG	4550
4551	ACGGCGCCGG	CGCCGGATGA	GGCGGGGGCA	AAAAAGGCTG	ATCTGCAACT	4600
4601	GGTCAAGGGC	GAAAGCATCA	CGATCCCCGT	CATCAATGAT	GGTGGAGTGA	4650
4651	ACGGCTATTT	CCTCAGCCGG	ATTTGCTGTC	GCGTCGACAA	GGCCAAGATG	4700
4701	GCGAAGATCG	AACTGCCGGC	AACACAGCTG	ATGACCGATG	AACTGTTTAC	4750
4751	GCTTCTGGCG	GGTTCTTCCA	TGGTCAACAT	TGCCAATATC	TCCACCTTCG	4800
4801	ATCCCGAAGC	CTTCAAGCAG	CGCATCCGCG	AAGGGTTGAA	CAAGAGGCTC	4850
4851	GATGACGAGG	TCGTGAGGGA	CGTGCTGATC	GAGCAGCTCG	ATTACCTGTT	4900
4901	CGAAAGCCGA	CATCCGCGAG	CTTAAGGACA	ATGGTTCGCC	CCACTCTGTC	4950
4951	AAACTCGTCG	AGGGCGAAAA	AGCCGAAGCA	GGAGAGGCTG	CGGCACCCAG	5000
5001	CCACTGACGG	AGAGGGAGCG	GCGGCGAAAC	GGCACGCTCT	GGTTAATCAT	5050
5051	CTCTTGCTCC	ATTCACGAAC	ATTCGAGAGA	TTGCTTTCAA	GTAAATTCA	5100

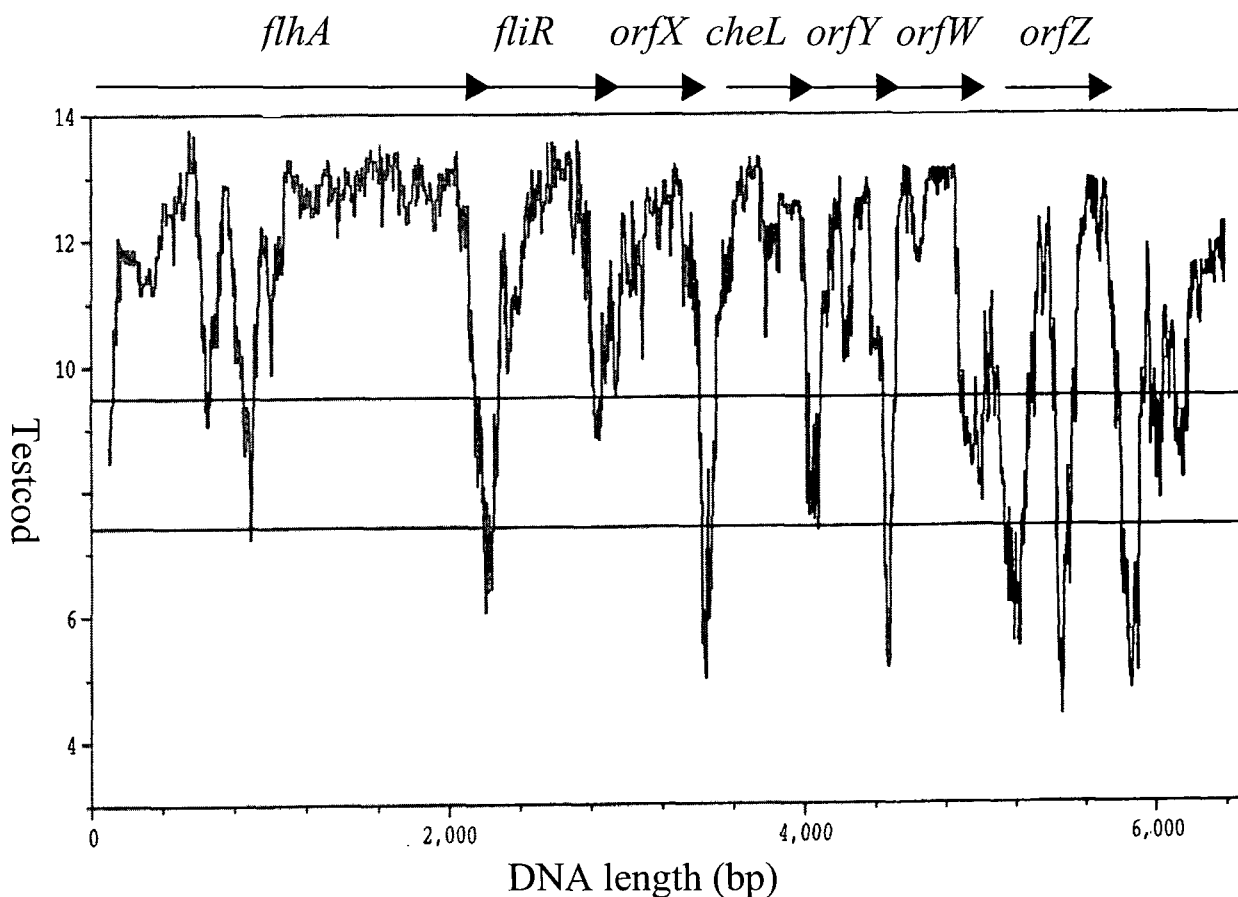


5101	ACCTGCACCT	GTATAACAAA	CTCATGAACC	ACCGCCCCGC	AGGCCATCCG	5150
5151	CCGGGCCGGT	GGCGATCGGT	TTGGCATTG	GGGTTCGGCA	ATGAATTTG	5200
5201	CGTCAGGTGT	TACGCGCCGG	GGATTGCAAC	GGAATATCCA	CGGGCTGCGC	5250
5251	GTCTGTGATC	TCGACTGGAA	TGGCGCGCTG	GAAGTGGTCC	GCGGCAAGGC	5300
5301	TTCGGCCTGT	GAAGGGCACA	CGATGCTCTC	CTTCCTGACC	ATCGAGAAGA	5350
5351	CAAAGTTGTC	ACTCAAGGAT	AATGCCTATC	GGCAAATTCT	GGAAAGCTGC	5400
5401	CTGCTTCTGC	CGCAGGGCAA	GGGCATGAAA	GCCGCCGCGC	GGGCCGAACA	5450
5451	TGGCGAGCCT	CTGCCCGCCA	CCATCGATGG	CGTTGCCTTC	ACCATGGCGC	5500
5501	TTCTGACCTA	CATGGCTGTG	CCGAAACGCA	TCGGCGTGGC	GGGCGACGAT	5550
5551	GCGGCGCAGA	TCGGTGAAGT	GCTTGCCCGG	CTGCGTGCCC	ATGCGCCATG	5600
5601	GCAAGACTTC	GTCATGCTGA	ACCGGGACGC	GGCCGGCGTG	AAGGTGGATG	5650
5651	TGCTCCTGGC	AGGAATGCTC	AAGGAAAATC	AGGAGACATG	GCTGCACCGC	5700
5701	AGCGTCAGCC	GTGAGGATGC	GCGTGTACAC	ATTGCAGTCG	GGCCCCTGTT	5750
5751	CAAGGTCCTG	TCCTCAGAAG	TGGCGGGTAT	GCCGGACATT	TTCCGCAAGC	5800
5801	TGCATATGAG	CTGGCTCTAC	AGCCTTTGTG	CCGAACCCTG	GCACATCGCG	5850
5851	CTCGGCAAGA	GCTGATATC				5869



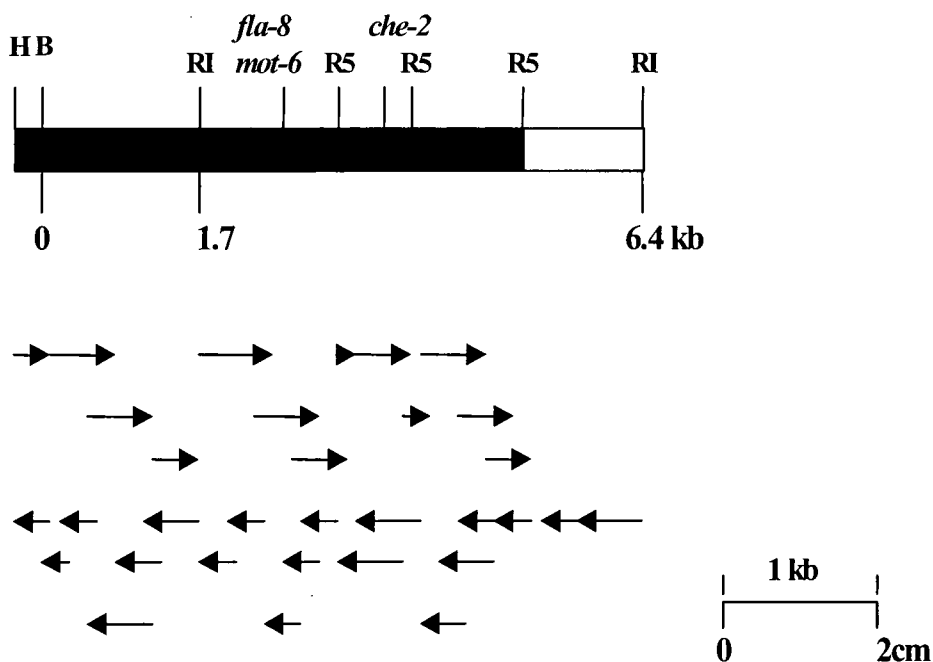
**Figure 3.4.3 DNA Strider open reading frame map of the sequenced region of pDUB1905 showing the location and name given to each identified open reading frame.**

Numbers on the vertical axis indicate the three different sense (1, 2 & 3) and antisense (-1, -2 & -3) reading frames. For each reading frame, short vertical lines indicate possible start sites (ATG or GTG), full length vertical lines indicate possible stop sites (TGA, TAA or TAG).



**Figure 3.4.4 Diagram of Testcode result of the sequenced region of pDUB1905.**

Regions above the upper horizontal line (9.4) are considered as coding regions with likelihood greater than 95%. Conversely the regions below the lower horizontal lines (7.4) indicate with 95% confidence as non-coding regions of DNA by Testcode. Area between these two values (9.4 and 7.40), are considered as “window of vulnerability” which Testcode can not predict significantly. The horizontal arrows above the diagram indicate the genes and the *orf* found in this project.



**Figure 3.4.5 Strategy for the partial sequencing of pDUB1905**

The dark box in this figure represents only the sequenced region of pDUB1905 and a DNA fragment (H-B in the diagram) synthesised by PCR. The letters in bold above the box is showing the position of selected restriction enzyme sites. The position corresponding to the mutants, *fla-8*, *mot-6* and *che-2*, have also shown above the box. The horizontal arrows below the box represent plasmids generated to sequence the region.

region of the cosmid 1905 is shown in figure 3.4.2. The open reading frame map and TESTCODE of this alignment are shown in figure 3.4.3 and 3.4.4. The strategy applied to fully sequence the alignment is shown in figure 3.4.5. Homology searches for possible motility or chemotaxis genes revealed that the alignment had similarity to different genes including *flhA*, *fliR*, and *cheL* in various bacteria. Some characteristics of the potential genes and *orfs* found in this work have been shown in table 3.4.1.

### 3.5 The *flhA* homologue:

Approximately 100 bp downstream of the 5' end of the alignment an open reading frame starts which shows to have a significant homology at the DNA and protein levels, to *flhA* gene of many bacteria containing both Gram-negative and Gram-positive bacteria. This open reading frame that finishes at position 2214 bp, is 2108 bp long and its protein product contains 703 amino acids, with predicted molecular weight of 76.86 kDa. Figure 3.5.1 shows the sequence of the *flhA* homologue along with the predicted gene product. The range of identity and similarity of homology between *Agrobacterium flhA* and the corresponding homologues in other genes were 40-60% and the similarity was 54-69% over at least 679 amino acids. FlhA is a member of the protein delivery system called the flagellum-specific export pathway/apparatus (see section 1.4.3).

The *Agrobacterium flhA* product was also found to have sequence similarity to the virulence proteins that are involved in the typeIII secretion protein apparatus in several bacteria, including *Yersinia pestis* and *Y. enterocolitica* *LcrD*, *Shigella flexneri* *MxiA* (*VirH*), *Salmonella typhimurium* *InvA* and *SsaV*, *Pseudomonas aeruginosa* *PcrD*, and *Y4yR* a probable translocation protein from *Sinorhizobium sp. NGR234*. Table 3.5.1 shows the proteins with sequence homology to FlhA. A multiple alignment was carried out using CLUSTALV at HGMP, the results of which are shown in figure 3.5.2. As can be seen, the sequence similarity between the ORF and the homologues is throughout their entire lengths, however it does not usually cover the first 32 amino acids.

A potential ribosome-binding site (rbs), that bears a weak similarity to the Shine-Dalgarno sequence, was found 19 bp upstream of the *flhA*. There are also nucleotides that have similarity to the consensus ribosome binding site sequence 5 bp

after the start codon of the *orf* and 7 bp upstream of the second methionine codon (M). The size of the FlhA protein varies in different bacteria and is between 673 amino acids to 733 amino acids. In most cases the FlhA in other bacteria is shorter than the protein product of this supposed *flhA*. Therefore it is possible to speculate that the gene may start from the second M codon.

A possible transcription terminator, with a stem-loop motif, was identified upstream of *flhA* using the Macintosh program DNA Strider and the terminator program at SEQNET. The putative terminator consists of a stem with 8 bp and a loop of 3 bp. Finding a terminator in this region suggests that this sequence could be the first gene of an operon. *flhA* is often the middle gene of an operon starting with *flhB* as in *C. crescentus*, *S. typhimurium*, *Y. enterocolitica* or *che/fla* in *Bacillus subtilis* (123, 172). Attempts to identify a promoter upstream of *flhA* were unsuccessful. Therefore, the translated sequence upstream of *flhA* was searched for significant homology to any other flagellar genes. No similarity was found in this region. This might be because the sequence, which contains only 106 bp is too short to reveal any homology with relevant genes, or there is a gap between *flhA* and the previous gene, as can be seen in *B. subtilis* (37) where a gap of 32 bp exists between *flhB* and *flhA*.

The hydropathy profile of the *A. tumefaciens* FlhA homologue, using the Kyte and Doolittle algorithm, revealed a similar pattern to that of other known homologues. The resemblance of hydropathy pattern between *A. tumefaciens* FlhA and *E. coli* FlhA is less than FlhA from *C. crescentus*. They have seven membrane-spanning segments, while *E. coli* has only four transmembrane domains. The result showed that the amino acid sequence could be divided into two distinct domains. The amino-terminal half of the FlhA protein contains seven strongly hydrophobic regions, predicted to be membrane-spanning segments, suggesting that this protein could be an integral membrane protein. There was no cleavable signal sequence. In contrast, the carboxyl-terminal is quite hydrophilic and is likely to represent a water-soluble domain. Figure 3.5.3 shows the hydropathy profile of FlhA and four of its homologues from flagellum-specific export and type III export systems. The predicted transmembrane domains of the FlhA, using SOSUI program at HGMP is shown in table 3.5.2.



**Figure 3.5.1** The entire coding and untranslated regions of *flhA* homologue and the predicted gene product.

The *flhA* coding regions begins at nucleotide position 106 and finishes at 2214. The possible start and stop codons are overlined and labelled *flhA* and "\*" respectively. The start codon of next homologue, which overlaps *flhA* stop codon, giving an "ATGA" configuration, is also overlined and marked *fliR*.

*HindIII* (at position 1 bp) and *BamHI* (at position 308 bp) denote the beginning of the sequence (and also PCR product) and beginning of cosmid pDUB1905 respectively. Two possible ribosome-binding sites, 19 bp before and 10 bp after the start codon, are overlined and labelled "rbs". The potential stem-loop structure, which may be a transcription termination signal is also shown, by staggered arrows and marked "TM".

HindIII

1 aag ctt cag cgg cta ggg ata ggc tga agg cat cgg ccc gaa ttg cgt tgc ctt ttc gaa 60

rbs

61 aac cga tgc ccg gaa ccg tgc gac ggc ttt tcc gtc gcc ctt ctc ATG AAG AGA CAG GAC 120

rbs

1 -----> TM <----- M K R Q D 5

flhA

121 GAA TTC ATG GCG CAG CCA CCA GTC ATC TCC TTG CCC AAG GTC AGT CCG AGC ATG CGG GAT 180

6 E F M A Q P P V I S L P K V S P S M R D 25

181 ATC GGT TTC GCA TTG GGC ATC ATT TCG ATC CTT TGC GTT CTG TTT CCG CCT ATT CCG GTC 240

26 I G F A L G I I S I L C V L F P P I P V 45

241 ATG CTG GTG GAC ATC GGT CTG GCG TTC TCG ATT ACG CTT TCG GTC CTC ATC CTC ATG GTG 300

46 M L V D I G L A F S I T L S V L I L M V 65

BamHI

301 GCG CTC TGG ATC CAG CGC CCG CTG GAC TTC TCG TCA TTC CCG ACC GTG CTG CTG ATT GCC 360

66 A L W I Q R P L D F S S F P T V L L I A 85

361 ACA ATG ATC CGT CTG TCG CTG AAC ATT GCG ACG ACG CGT GTC ATC CTT TCG CAT GGT AAC 420

86 T M I R L S L N I A T T R V I L S H G N 105

421 GAG GGG CCG ACG GCG GCA GGT GGC GTG ATC GCG GGT TTC TCC AGC CTC GTC ATG TCC GGT 480

106 E G P T A A G G V I A G F S S L V M S G 125

481 GAC TTC GTC ATC GGT CTC ATC GTC TTC CTG ATC CTG ATC ACG GTC AAC TTC ATC GTC ATC 540

126 D F V I G L I V F L I L I T V N F I V I 145

541 ACC AAG GGT GCG ACG CGT ATC GCC GAA GTC GGC GCG CGC TTC ACC CTA GAC GCC ATC CCC 600  
146 T K G A T R I A E V G A R F T L D A I P 165  
  
601 GGC AAG CAG ATG TCG ATC GAT GCG GAT CTG TCC GCC GGC ATC ATC GAT GAA AAG GAA GCG 660  
166 G K Q M S I D A D L S A G I I D E K E A 185  
  
661 CAG CGT CGC CGC CGG GAA CTG GAA GAA GAA AGC TCT TTC TTC GGT TCG ATG GAC GGT GCC 720  
186 Q R R R R E L E E E S S F F G S M D G A 205  
  
721 TCG AAA TTC GTC CGC GGC GAC GCG ATC GCC GGC CTC ATC ATC ACC GCG ATC AAC GTT TTC 780  
206 S K F V R G D A I A G L I I T A I N V F 225  
  
781 GGC GGC ATC ATC ATC GGT TAT TTC CGC CAC GGC ATG CCG ATC GGC GAA GCG GCC GAT GTT 840  
226 G G I I I G Y F R H G M P I G E A A D V 245  
  
841 TTC GTC AAG CTT TCG GTC GGT GAT GGT ATC GTC TCG CAG ATC CCG GCC CTG ATC GTA TCG 900  
246 F V K L S V G D G I V S Q I P A L I V S 265  
  
901 CTG GCG GCC GGC CTT CTG GTG TCG CGC GGT GGC ACA GCC GGA TCG ACG GAT CAG GCC GTC 960  
266 L A A G L L V S R G G T A G S T D Q A V 285  
  
961 ATC AAT CAG CTG AGC GGT TAT CCC CGC GCT TTG ATG GTC TCG GCC ATG CTG ATG GGT CTG 1020  
286 I N Q L S G Y P R A L M V S A M L M G L 305  
  
1021 CTC GCC GTC ATT CCG GGC CTT CCC TTC GTG CCC TTC ATT TTC CTC GGC GGC ATC ATG GCC 1080  
306 L A V I P G L P F V P F I F L G G I M A 325  
  
1081 TTT GGC AGC TGG TAT ATT CCG CGC CAG GCC GAA GCC GAA AGC GCG CTT CGC CGC CAG GAA 1140  
326 F G S W Y I P R Q A E A E S A L R R Q E 345  
  
1141 GAG GAA AAC AAG GTT CTC CAG ACC ACC GAG GCG GAA AAG GAT TCC GTC AAG CAG GTG CTG 1200  
346 E E N K V L Q T T E A E K D S V K Q V L 365  
  
1201 AAG ACC GCC GAG ATC GAG CTC GCG CTC GGC AAG CAG GTT TCC ACC CGC CTT CTC GGC GCG 1260  
366 K T A E I E L A L G K Q V S T R L L G A 385  
  
1261 CAT CAG GAA CTG GCC TTC CGC GTC GGC AAG ATG CGC AAG AAA TTC GCC ACG CAA TAT GGC 1320  
386 H Q E L A F R V G K M R K K F A T Q Y G 405  
  
1321 TTT GTG GTG CCG GAG ATC AAG GTC TCC GAC GAC ATC ATG ATC CCG GAA AAG GCC TAT CAG 1380  
406 F V V P E I K V S D D I M I P E K A Y Q 425  
  
1381 ATC CGC GTC CAC GGC ACG ACG ATC GCG TCC AGC AAC CTG CGC GTC GGC GAT GTG CTG GTC 1440  
426 I R V H G T T I A S S N L R V G D V L V 445  
  
1441 GTG ACG GGG GCA GGG CGC AAG CCG AGC ATC CCA GGT GAT GAA ATC CGC GAA CCC GCT TTC 1500  
446 V T G A G R K P S I P G D E I R E P A F 465



1501 GGC ATG CCG GCC GTC TCG ATC CTC GAG ACC TTT ACC GAG GAT CTG AAG CGC GAA GGC TTC 1560  
466 G M P A V S I L E T F T E D L K R E G F 485

1561 CAC CCG ATC GAT AAC GTC TCT GTC GTG CTG ACG CAT CTG AGC GAG GTT ATC CGC AAC AAT 1620  
486 H P I D N V S V V L T H L S E V I R N N 505

1621 CTG CCG CAG CTT CTG TCC TAC AAG GAC GTC AAG ATC CTC ATC GAC AGG CTT GAT CCC GAA 1680  
506 L P Q L L S Y K D V K I L I D R L D P E 525

1681 TAC AAG AAG CTG GCT GAC GAG ATT TGT TCG TCG CAC ATG TCC TAT TCC GGC CTG CAG GCG 1740  
526 Y K K L A D E I C S S H M S Y S G L Q A 545

1741 GTG CTG AAG TTG CTG CTG GCC GAG CGC GTA TCG ATC CGC AAC CTG CAT CTC ATT TTG GAA 1800  
546 V L K L L L A E R V S I R N L H L I L E 565

1801 GCG GTG GCA GAA CTC GCA CCG CAT GTG CGC AAG ACC GAG CAG ATC GTC GAA CAT GTC CGC 1860  
566 A V A E L A P H V R K T E Q I V E H V R 585

1861 GTT CGC ATG TCG CAG CAG CTC TGC GGC GAC CTT GCC GAC AAT GGC GTG CTG CGC GTC CTG 1920  
586 V R M S Q Q L C G D L A D N G V L R V L 605

1921 CGA CTC GGC AAC AAG TGG GAC ATG GTC TTC CAT CAG GCG CTG AAA CGC GAC CAG AAG GGC 1980  
606 R L G N K W D M V F H Q A L K R D Q K G 625

1981 GAA ATC GTC GAA TTT GAC ATC GAT CCG CGC CAT CTC GAG GAG TTT TCC GAG CAG GCA TCG 2040  
626 E I V E F D I D P R H L E E F S E Q A S 645

2041 AAA GTT ATC CGT GAA TTC ATG GAT CGC GGG CTG CCC TTT GTC CTT GTA ACC TCG CCG GAA 2100  
646 K V I R E F M D R G L P F V L V T S P E 665

2101 ACG CGG TCC TAT GTG CGC ATG ATT ATC GAG CGA CTC TTT GCG ACC CTG CCG GTT CTT TCA 2160  
666 T R S Y V R M I I E R L F A T L P V L S 685

*fliR*

2161 CAT GTG GAA CTG GCC AAG GGT CTC GAG ATC AAG ATT CTG GGC GCC ATT TCA TGA 2214  
686 H V E L A K G L E I K I L G A I S \* 703

**Table 3.5.1 A few of the proteins with sequence homology to potential *A. tumefaciens* FlhA**

Organism	Protein	Identity (%)	Similarity/Overlap (%/bp)	Length of protein	function
<i>C. crescentus</i>	FlhA	40	60/679	700	flagellar biogenesis
<i>Aquifex aeolicus</i>	FlhA	36	56/682	678	flagellar biogenesis
<i>Proteus mirabilis</i>	FlhA	36	57/677	696	flagellar biogenesis
<i>Eurasian enterocolitica</i>	FlhA	34	56/677	692	flagellar biogenesis
<i>Vibrio parahaemolyticus</i>	FlhA	33	56/707	699	Flagellar biogenesis
<i>B. subtilis</i>	FlhA	35	55/684	677	Flagellar biogenesis
<i>S. typhimurium</i>	FlhA	35	56/679	692	Flagellar biogenesis
<i>Escherichia coli</i>	FlhA	34	55/679	692	Flagellar biogenesis
<i>Thermotoga maritima</i>	FlhA	34	54/681	678	Flagellar biogenesis
<i>Helicobacter pylori</i>	FlhA	33	54/710	733	Flagellar biogenesis
<i>Compylobacter jejuni</i>	FlhA	31	52/728	724	Flagellar biogenesis
<i>Zymomonas mobilis</i>	FlhA	30	53/676	707	Flagellar biogenesis
<i>Borrelia burgdorferi</i>	FlhA	29	54/689	697	Flagellar biogenesis
<i>Treponema pallidum</i>	FlhA	30	50/686	707	Flagellar biogenesis
<i>Salmonella typhimurium</i>	InvA	31	51/682	685	Invasion protein
<i>Shigella flexneri</i>	MxiA	30	53/679	686	Virulence protein
<i>Yersinia entocolitica</i>	LcrD	30	50/705	704	Low-calcium response protein
<i>Yersinia pestis</i>	LcrD	29	52/702	704	Low-calcium response protein
<i>Pseudomonas syringae</i>	HrpI	28	50/689	695	Hypersensitivity response protein
<i>Pseudomonas aeruginosa</i>	PcrD	29	51/702	706	TypeIII secretion protein
<i>Sinorhizobium sp. NGR234</i>	Y4yR	30	54/694	697	TypeIII secretion protein

**Table 3.5.2 Transmembrane domain of *A. tumefaciens* FlhA.**

Transmembrane NO.	N terminal	Transmembrane region	C terminal	length
1	18	KVSPSMRDIGFALGIISILCVLF	40	23
2	48	VDIGLAFSITLSVLILMVALWIQ	70	23
3	78	FPTVLLIATMIRLSLNIATTRVI	100	23
4	119	SSLVMSGDFVIGLIVFLILITVN	141	23
5	211	GDAIAGLIITAINVFGGIIIGYF	233	23
6	253	DGIVSQIPALIVSLAAGLLVSRG	275	23
7	298	VSAMLMGLLAVIPGLPFVPFIFL	320	23

Figure 3.5.2 Multiple alignment (pretty box) of the potential *A. tumefaciens* FlhA to those of flagellum-specific export component and type III export proteins

Abbreviation used for bacteria are as follow: AGRO= *A. tumefaciens*, ECOL= *E. coli*, BACSU= *Bacillus subtilis*, YERPE= *Yersinia pestis*, YEREN= *Y. enterocolitica*, SHIF= *Shigella flexneri*.

FLHA-AGRO	-----MKRQDEFMAQPPVVISLPKVSPSM
FLHA-ECOL	-----MSNLA AHLRLPANLKSTQW
FLHA-BACSU	-----MST
LCRD-YERPE	-----MNP HDLEULNRIGER
INVA-YEREN	MYSFFNTLTVTKIISRLILSIGLIFGIFTYGFSSQNYFNSEALENPAEHNEAFNKIISITG
MXIA-SHIF	-----MIQSFLKQVSVTK
FLHA-AGRO	RDIGFALGITSILCVLPFPPIPVRLVDIGLAFSITLSVLILMVALLWIORPLDFS-SFPTVL
FLHA-ECOL	QILAGPILILLILSMMVLPPLPAFILDLELFTFNIALSIVMLLVAMFTORTLEFA-AFPTIL
FLHA-BACSU	RDL SVLISVVLIVAIIVVIFFPPLLSILIIINISLALIVLLTTMNMQEAQCFS-IFPSLL
LCRD-YERPE	KDITMLAVLLLA VVFMVLPPLPLVLDILIAVNMITSVVLMLAIYINSPLQFS-AFPAVL
INVA-YERCA	TSLAVSGNASNITRSMVNDAAANQEVKHWLNRFGITQVNVNFDKKFSLKESSLDWLLPWFYD
MXIA-SHIF	PELIILVLMVMITAMLIIPLETYLVDVFLIGLNLVLAALWFMGSFYIERILSFS-TFSSVL
FLHA-AGRO	L IATMI--RLSLNIATTRVILSHGNEGPTAAGGVIAGFSSLW---MSGDFVIGLIVFLI
FLHA-ECOL	LFTTLL--RLALNVASTRIILMECHTGAAAAGKVVEAFGHFL---VGGMFAIGIVVFI
FLHA-BACSU	L LTLF--RLGLNVSTTRSILSHGE---CGKVVEFTGNFV---VGGMVLVGLVVFII
LCRD-YERPE	LVTTLF--RLALSWSTTRMILLQAD---AGQIVYTFGNFV---VGGMLIVGLIVTFEI
INVA-YEREN	SASVVFFSQLGIRNKDSRNTILNIGAGVRTFOQSVMYGFNTFYDNDMTGHNHRIGVGAEAW
MXIA-SHIF	LITTLF--RLALSISTSRLLILVDAD---AGKIIITFGQFV---IGDSLAVGFVIFSI
FLHA-AGRO	LITVNFIVITKGATRIAEVGARFT-LDAMP GKQMSIDADLSAGIIDEKEAQRRELEEE
FLHA-ECOL	LVIINFMVITKGAGRIAEVGARFV-LDGM PGKQMAIDADLNAGLIGEDEAKKRSEVTOE
FLHA-BACSU	LIIHQFIVITKGAERVSEVAARFT-LDAMP GKQMSIDADLNAGMITEQEAHRREKVARE
LCRD-YERPE	ITIVQFLVITKCSERVAEVSARFS-LDAMP GKQMSIDGDMRAGVIDVNEARERRATIEKE
INVA-YEREN	TDYLQLSANGYFRLNGWHQSRDFVDYNERPASGGDIHVKA YLPALPOLGGKLKYEQYRGE
MXIA-SHIF	VTIVQFIVITKCSERVAEVAARFS-LDGM PGKQMSIDADLKAGIIDAAGAKERRSILERE
FLHA-AGRO	SSFE GSDMGASKFVRGDAIAGLIITAINVFGGI-----IIC YFRHGMPIGEAAD
FLHA-ECOL	ADFYGSDMGASKFVRGDAIAGLIMVINIVGGL-----LVGV LQHGMSMGHAAE
FLHA-BACSU	ADFYGAMDGASKFVKGDAIAGLIIVMINIIFGI-----VICMLQQGMSIOEAA S
LCRD-YERPE	SQMFGSDMGAMKFKVGDAIAGLIIFVNILGCV-----TICVTCKGLAAAEALQ
INVA-YEREN	RVALFGKDNLSNPYAVTTGLIYTPIPFITLGV DQRNGKS RQHEIQUNLQMDYRLGESFR
MXIA-SHIF	SQLYGSFDGAMKFIKGD A IAGLIIFVNILGGI-----SVGMSQHGM SLSGALS
FLHA-AGRO	VFVKLSVGDG-----IVSQIPALIVSLAAGLLVSRGCTA-----GSTDQA
FLHA-ECOL	SYTLTITGDG-----LVAQIPALVISTAAGVIVTRVSTD-----QDVGEQ
FLHA-BACSU	HFTMLTVGDG-----IVSQIPALLISTATGIVVTRAASE-----GNLGH D
LCRD-YERPE	LYSILTVGDG-----MVSQVPALLIATAGIIVTRVSSD-----SSDLGSD
INVA-YEREN	SQFSPA VVAGTRLLAESRYNLVERNPNIVLEYQKONTIKLAFSPA VLSGLPGQVYSVSAQ
MXIA-SHIF	TYTILTITGDG-----LVSQIPALLISTAGFIIVTRVNGD-----SDNMGRN
FLHA-AGRO	VINQLSGYPRA-----LMVSAAML MGLLAVIPGL-BFVPPFI FLGGIMA-----FG
FLHA-ECOL	MVNQLFSNP SV-----MLLSAAVILGLLGLVPGMPNLVFLIFTAGILG-----LA
FLHA-BACSU	ITGOLFAYPKL-----LYVAAATIMLLGIFTPI-GILLTGPLAGLLA-----FG
LCRD-YERPE	IGKQVVAQPKA-----MLTGGVLLLLFGLIPGF-PTVTFILALLLVG-----CG
INVA-YEREN	IQSCSALQRIILWDDAQWIAVGGKLIPVSA TNYNVVLPPYKEMAPASRTVGKTGESEAAVN
MXIA-SHIF	IMSGIFGNPFV-----LI VTSALALAIGMLPGFPFFVFFLI AVTLTA-----LF

FLHA-AGRO SWYIPRQAEAESALRRQEEENKVLQT-----TEAEKDSVKQV-L-KTAEIELALGKQVS  
FLHA-ECOL WUIRGREQKAPAEPKPKVKMAENNTV-----VEATWNDVQLE--DSLGMVEVGYR-LIP  
FLHA-BACSU AYTLSSKSGKEKEEVDEILEEEAEVDE-----LKSPESVVQLLHI-DPIEFEEFGYG-LIP  
LCRD-YERPE GYMLSRKQSRNDEANQDLQSILTS GSGAPAARTKAKTSGANKGRIDGEQEAFAFAMTVPLLLID  
INVA-YEREN TYTLSATIDNHNSSNPATLTVTIVQQPQFVITSEVTDDGALADGRTPITVKFTVTINIDG  
MXIA-SHIF YKKVVEKEKSLSESDSSGYTGTFDI-----DNSHSSSLAMIENTLDAISSSETVPLILLF

FLHA-AGRO TRLLGAH-----QELAFRVGK--MRKKFATQY-GFVVPEIKVSDDIMPEK--AYCI  
FLHA-ECOL MVDFOCD-----GELLGRIRS--IRKKFAQEM-GFLPPVWHIRDNMDLCPA--RYRI  
FLHA-BACSU LADANCG-----GDLLDRIVM--IRRQLALEL-GLVFPVVRIRDNIALCPN--EYRL  
LCRD-YERPE VDS SQCEAL--EANALNDELVR--VRRALYLDL-GVPFPGIHLRFNEGMEG--EYII  
INVA-YEREN TPVAEQEGVITTSNGALPSKVTKKTDAGQVISIALTSFTVGGSVVTLDIQQQQTVDVRF  
MXIA-SHIF AENKINA-----NDNEGLIER--IRSQFFIDY-GVRLPTILYRTSNELKVD--DIVL

FLHA-AGRO RVHGTTIASNLRVGDVLVVTGAGRKPS--IPGDEIREPAFGMPAV-----SILETTTE  
FLHA-ECOL LMKGVEIGSGDAYPGRULAINPGTAAGT--LPGEATVDPAFGLNAI-----WIESALKE  
FLHA-BACSU KIKGNEVAKGELLLDHYLAMSPTPEDDL--IEGIETVEPSFGLPAK-----WISEAVKD  
LCRD-YERPE SLOEVPVARGELKAGYLLVRESVSQLELLG-IPYEKGEHLLPDQEA--WVSVEYEE  
INVA-YEREN AVLPPDVTNSSFNMSPSDIVADGSMQSILTFMPRNKNEFVSGITDLEFIQSGVPVTISP  
MXIA-SHIF LLINEWRADSFNTYFDKVCITDENGDIDALG-IPVVSISYNERVITS-----WVDVSYTE

FLHA-AGRO DLKREG--FHPIDNVSVVLTHLSEVIRNNLPQLLSYKDVKILIDRI-DPEYKKLADEIC  
FLHA-ECOL QAQIQG--YTVVEASTVWATHLNHLISQHAELFCRQEAQQLLDV--AQEMPKLTEDIVP  
FLHA-BACSU EADM LG--YTVWDPASVSTHITTEKIKQHAEHLIGROET-----  
LCRD-YERPE RLEKSO--LEFFSHSQVLTWHLSHVLREYAEFIGIQETRYLLEQM-EGGYGELIKEV-Q  
INVA-YEREN VTENADNYTASVVGNSVGDVDITPQVGGESLDLLOKRIITLYPIPKITSIAVNGEQFATDK  
MXIA-SHIF NLTNID--AKIKSAQDEFYHQLSQALLNNINELFGIQETKNMDDQF-ENRYPDLLKEV-F

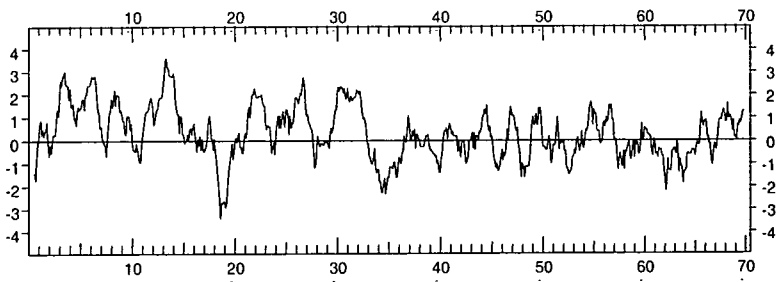
FLHA-AGRO SHMSYSGLOAVLKLLEAERVSIRNLHLILEAVAEAPHVRKTEQIVEHVRVRMSQOLCGD  
FLHA-ECOL GVWTLTTLHKVLCNLLDEKWPIRDRTILETLAEHAPIQS DPHELTAVVRVALGRAITQQ  
FLHA-BACSU -----  
LCRD-YERPE RIWPLQRMTEILQRLVGEDISIRNMRSILEAMVVEWGOKEKDVVQLTEYIRSSLKRYICYK  
INVA-YEREN GFPKTTFNKATFCLVMNDDVANNTQYDWTSSYMAASAFVDNQGVNIAYKTYGSTVTVTAK  
MXIA-SHIF RHVTTIQRTISEVLCRLLGENISVRNLKLIMESLALWAFREKDVITLVEHVRASLSRYICSK

FLHA-AGRO -LADNGVLRVLRGLGNKMDMVFHQALKRDQKGEIMEFDIDPRHLEEFSEQASKVIREFMDR  
FLHA-ECOL WFP GKDEMHVIGLDTPLERLLLOALOGGG-----GLEPGLADRLLAQTQEAISRQEM  
FLHA-BACSU -----  
LCRD-YERPE YANGNNILPAYLFDQEEKIRSGVRQTSAGSYL--ALEPAVTESSLEQVRKTIQDLSQI  
INVA-YEREN SKKFPSYTATYQFKPNLWVFSGTMSLQSSIEASR--NCORTDFTALIESARASNGSRSPD  
MXIA-SHIF -IAVSGEIKVVMLSGYIEDATRKGIRQTS GGSFL--NMDIEVSDEVNETLAHALRELNA

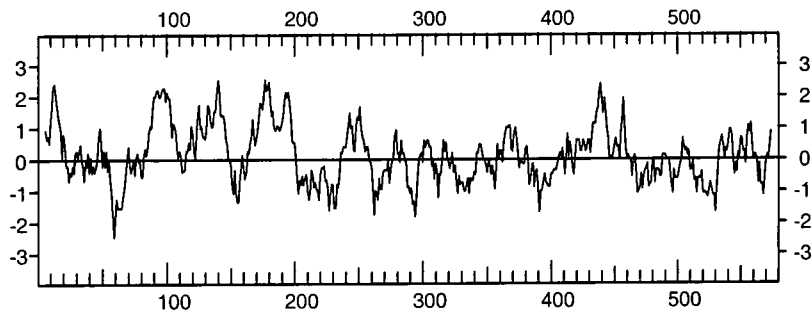
FLHA-AGRO GL--PFVLVITSPETR SYVRMIERLEATLPVLSHVELAKGLEIKILGALS-----  
FLHA-ECOL GA--PFVLLVNHALLRPLLRFRLRRSLPOLVVLNLELSDNRHIRMTATIGGK-----  
FLHA-BACSU -----  
LCRD-YERPE QS--KPVLIVSMDIRRYVRKLI ESEYYGLPVLSYQELTQQINIQPLGRICL-----  
INVA-YEREN GTLWGEUGSLATYDSA EUPSGNYWTKKTS TDFVTMDMTTGAIPTSAATAYPLCAEPQ  
MXIA-SHIF KK--NFVLLVSVDIRRFVKRLTDNRFKSLVVISYAEIDEAYTINVVKTI-----

**Figure3.5.3** Hydropathy profiles of the *A. tumefaciens* FlhA homologue and FlhA from other bacteria, LcrD and MxiA from typeIII export system.

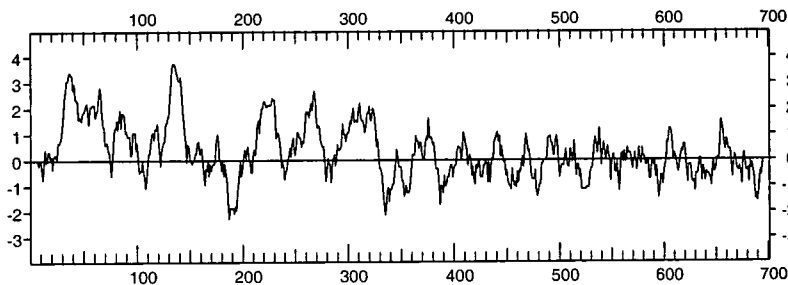
**FlhA from**  
*A. tumefaciens*



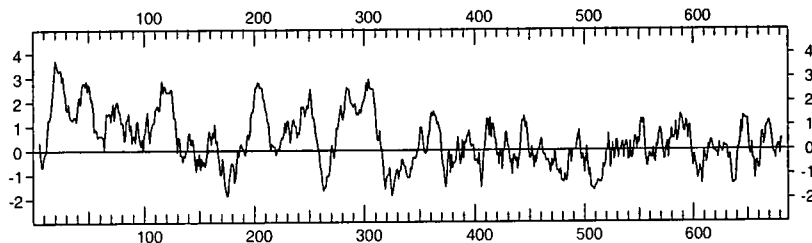
**FlhA from**  
*E. coli*



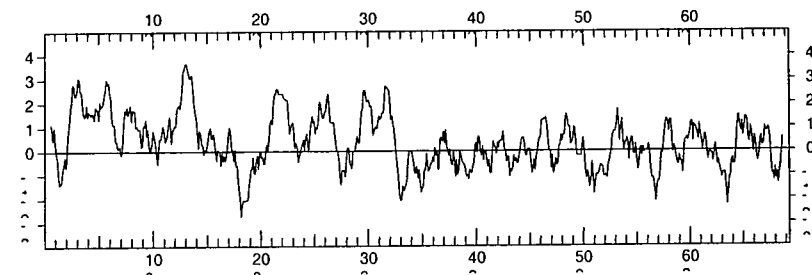
**FlhA from**  
*C. crescentus*



**LcrD from**  
*Yersinia pestis*



**MxiA from**  
*Shigella flexneri*



### 3.6 The *fliR* homologue:

There is an open reading frame starting immediately downstream of *flhA*, at position 2211 and finishing at position 2966. The open reading frame spans a region of chromosomal DNA which contains two behavioural mutants *fla-8* and *mot-6*. These two mutants, which were generated by insertion of Tn5 transposon, share an identical site. The putative start codon of this *orf*, which has 756 bp (252 amino acids), with predicted a molecular weight of 26.87 kDa, overlaps with the stop codon of *flhA* by 3 bp, to give an "ATGA" configuration.

Two putative ribosome-binding sequences (TGGG and AGGG) resembling the Shine-Dalgarno sequence were identified 9 and 30 bp upstream of the *fliR* homologue respectively. The latter is unlikely to be a ribosome-binding site as it is too far from the start codon. There was no terminator-like sequence downstream of the *fliR* stop codon. The DNA sequence of the *fliR* homologue along with the predicted gene product and potential ribosome-binding sites are shown in figure 3.6.1.

The product of the open reading frame, which is transcribed in the same direction as *flhA*, showed sequence similarity with FliR of several bacteria, at the protein level. It was, therefore given the name *fliR*. The FliR protein is a flagellar biosynthetic protein, and like FlhA is a member of the flagellum-specific export proteins that are involved in transport of extracellular flagellar proteins. A search of the GenBank database showed the *A. tumefaciens* FliR to have significant homology to the *Erwinia carotovora* MopE, *Erwinia herbicola* and *Erwinia amyloyora* HrcT, *Shigella flexneri* and *S. sonnei* SpaR (Spa29), *Chlamydia pneumonia* YopT, *Yersinia enterocolitica* YscT, *S. typhimurium* SsaT, *Sinorhizobium fredii* Y4yN, *Pseudomonas syringae* pv. *syringae* HrpX, all of which are involved in secretion of virulence factors into plant and animal cells. The range of identity between *A. tumefaciens* FliR and other homologue proteins was 20-29% and similarity was 38-51%, over at least 213 amino acids. The size of FliR proteins varies between 252 to 277 amino acids in different bacteria. Table 3.6.1 shows the proteins with sequence homology to the *Agrobacterium* FliR. A multiple alignment of the predicted *A. tumefaciens* FliR protein with their homologues was also carried out, which is shown in figure 3.6.2.

The *fliR* gene from *Caulobacter crescentus* is a member of *fliQR* operon which consists of only *fliQ* and *fliR* (238). The *E. coli* *fliR* protein is located in the *fliL* operon, and it is required for flagellar biogenesis (127). The *R. sphaeroides*

**Figure 3.6.1 The untranslated and coding regions of *fliR* homologue with its predicted translated product.**

The *fliR* homologue starts immediately after *flhA* at position 2211 bp and finishes at position 2966 bp. The start codon of the gene is underlined and labelled *fliR*, while the stop codon shown by "\*".

Two potential ribosome binding sites 9 bp and 30 bp upstream of the start codon are underlined and labelled "rbs" .

																				<b>rbs</b>	
2124	tat	cga	gcg	act	ctt	tgc	gac	cct	gcc	ggt	tct	ttc	aca	tgt	gga	act	ggc	caa	<u>ggg</u>	tct	2183
										<b>rbs</b>		<b><i>fliR</i></b>									
2184	cga	gat	caa	gat	<u>tct</u>	<u>ggg</u>	cgc	cat	ttc	<u>ATG</u>	ATA	ACC	GAC	CCG	CAG	GGA	ACA	ATC	ATT	GCA	2243
1										<b>M</b>	I	T	D	P	Q	G	T	I	I	A	11
2244	TTG	TTC	CTT	GCC	ATC	TGC	CGC	ATA	GGC	GCC	TGC	TTC	ATG	ACC	ATG	CCG	GGC	TTT	TCA	AGT	2303
12	L	F	L	A	I	C	R	I	G	A	C	F	M	T	M	P	G	F	S	S	31
2304	TCG	CGC	ATC	TCG	CCG	CAG	ATT	CGC	ATG	CTT	CTC	TGT	GTA	GCG	GTA	TCC	ATG	GCG	CTT	CTG	2363
32	S	R	I	S	P	Q	I	R	M	L	L	C	V	A	V	S	M	A	L	L	51
2364	CCC	GTC	CTG	TGG	GAC	ACC	ATC	TAC	CCA	AAG	GTT	TCC	GGT	GCG	AGC	CAG	GGC	GCG	GTC	GTT	2423
52	P	V	L	W	D	T	I	Y	P	K	V	S	G	A	S	Q	G	A	V	V	71
2424	GGC	CTC	ATC	TTC	TCC	GAG	GTT	CTC	ATC	GGC	GCA	ATG	TAT	GGG	CTG	ATC	GCC	CGT	TTC	TAT	2483
72	G	L	I	F	S	E	V	L	I	G	A	M	Y	G	L	I	A	R	F	Y	91
2484	ACG	CTC	GGT	TTC	CAG	TTC	ACC	GGT	GCG	CTC	ATC	GGC	GCT	TCC	ATC	GGC	CTC	AGT	GCG	CCC	2543
92	T	L	G	F	Q	F	T	G	A	L	I	G	A	S	I	G	L	S	A	P	101
2544	GGC	GGT	GCC	GAT	CCC	ATC	GAA	GAC	GTG	CAG	GAA	AAC	CAG	ATC	GCC	AAT	TTC	ATC	ACC	TTC	2603
102	G	G	A	D	P	I	E	D	V	Q	E	N	Q	I	A	N	F	I	T	F	123
2604	GGT	GGC	CTG	CTG	GTG	CTG	TTC	ATG	ATG	GAT	TTC	CAC	CAC	ATC	GTC	CTG	AAG	GCG	CTG	GTC	2663
122	G	G	L	L	V	L	F	M	M	D	F	H	H	I	V	L	K	A	L	V	143
2664	GAT	TCC	TAT	AGC	GCC	ACA	CCG	GTC	GGT	GCG	CTC	ATC	AGC	GGA	CAG	AAG	ATG	CTG	ATC	ACG	2623
144	D	S	Y	S	A	T	P	V	G	A	L	I	S	G	Q	K	M	L	I	T	163
2724	CTG	ACG	GAT	ACG	CTC	AGG	GCG	TCC	TTT	TCG	ATC	ATG	CTG	CGG	CTG	GCG	AGC	CCC	TTT	GTT	2683
164	L	T	D	T	L	R	A	S	F	S	I	M	L	R	L	A	S	P	F	V	183



NdeI

2784	ATT	TAC	GGC	<u>CAT</u>	<u>ATG</u>	TTC	AAC	GTC	GCG	GTC	GGT	CTC	ATC	AAC	AAG	CTG	GCG	CCG	CAA	ATC	2743
184	I	Y	G	H	M	F	N	V	A	V	G	L	I	N	K	L	A	P	Q	I	203
2844	CCG	GTC	TTC	TTC	ATA	TCG	ACA	CCC	TTC	GTT	CTG	GCG	GGT	GGT	CTT	TTC	ATG	CTT	TAT	TTG	2803
204	P	V	F	F	I	S	T	P	F	V	L	A	G	G	L	F	M	L	Y	L	223
2904	TCG	GTC	GCG	GGC	CTC	ATC	CGG	CAA	TTC	GTG	GAT	GGG	TTC	GGC	CCG	GTC	TTT	ATC	GGC	TTT	2863
224	S	V	A	G	L	I	R	Q	F	V	D	G	F	G	P	V	F	I	G	F	243
2964	TGA																				2966
244	*																				244

**Table3.6.1 Proteins with homology to potential *A. tumefaciens* FliR**

Organism	Protein	Identity (%)	Similarity /overlap (%/bp)	Length Of protein (aa)	Function
<i>Caulobacter crescentus</i>	FliR	29	51/231	251	Flagellar biosynthetic protein
<i>Pseudomonas putida</i>	FliR	25	49/244	258	Flagellar biosynthetic protein
<i>Escherichia coli</i>	FliR	26	44/217	261	Flagellar biosynthetic protein
<i>Erwinia carotovora</i>	FliR	27	43/219	261	Flagellar biosynthetic protein
<i>Rhodobacter sphaeroides</i>	FliR	25	40/228	269	Flagellar biosynthetic protein
<i>Salmonella typhimurium</i>	FliR	26	44/230	264	Flagellar biosynthetic protein
<i>Bacillus subtilis</i>	FliR	25	42/227	259	Flagellar biosynthetic protein
<i>Aquifex aeolicus</i>	FliR	26	43/250	258	Flagellar biosynthetic protein
<i>Borrelia burgdorferi</i>	FliR	25	40/178	269	Flagellar biosynthetic protein
<i>Vibrio parahaemolyticus</i>	FliR	23	42/231	260	Flagellar biosynthetic protein
<i>Helobacter pylori</i>	FliR	20	38/236	255	Flagellar biosynthetic protein
<i>Erwinia amylovora</i>	HrcT	26	43/230	265	TypeIII secretion protein
<i>Shigella flexneri</i>	SpaR*	22	44/214	256	Surface presentation of antigen
<i>Chlamydia trachomatis</i>	YopT	22	44/174	289	Translocation protein
<i>Salmonella typhimurium</i>	SpaR	23	42/205	263	Surface presentation of antigen
<i>Erwinia carotovora</i>	FliR**	27	42/172	180	TypeIII secretion protein
<i>Sinorhizobium sp. NGR234</i>	Y4yN	23	27/217	272	Translocation protein
<i>Yersinia pestis</i>	YopT	20	42/230	261	TypeIII secretion protein

\* formerly Spa29

\*\* formerly MopE (as mentioned in the text)

**Table3.6.2 Transmembrane domains of *A. tumefaciens* FliR.**

Transmembrane NO.	N terminal	Transmembrane region	C terminal	length
1	6	QGTIIALFLAICRIGACFMTMPG	28	23
2	36	PQIRMLLCVAVSMALLPVLWDTI	58	23
3	64	GASQGAVVGLIFSEVLIGAMYGL	86	23
4	120	FITFGGLLVLFMMDFHHIVLKAL	142	23
5	161	LITLTDTLRASFSIMLRLASPFV	183	23
6	211	PFVLAGGLFMLYLSVAGLIRQFV	233	23

**Figure 3.6.2 Multiple alignment (pretty box) of the potential *A. tumefaciens* Flir to those of flagellum-specific export system and type III secretion proteins.**

Abbreviation used for bacteria are as follow: AGRO= *A. tumefaciens*, ECOLI= *E. coli*, BACSU= *Bacillus subtilis*, ERWCA= *Erwinia carotova*, YEREN= *Yersinia enterocolitica*, SALTY= *S. typhomurium*.

```

FLIR-AGRO  ---M I T D P Q G T I I A L F L A I C R I G A C F M T M P G F S S R I S P Q I R M L L C W A V S M A L L P V L W D
FLIR-ECOLI M L Q V T S E Q W L S W L N Y F W P L L R V L A L I S T A P I L S E R S V P K R V K L G L A M M I T F A I A P S L P A
FLIR-BACSU  ---M N S I I D P A F L L V F I R I S A F F V T I P L E G H R M V P A V H R I G F A F F L A V I C F S T I D K
FLIR-ERWCA M S S D A I H T V Y Q F L F A L T L G A A R I Y P C L I L M P V F S F N I L K G M V R I G V V L A L S L M P A I G L C A
MOPE-ERWCA M L T F N S W D M V N W W S Q F F W P F V R I L A L I S T A P V F N E R A I G N R V K I G L G V L I T L L V A P Y L P L
YOPT-YEREN M S S D A I H T V Y Q F L F A L T L G A A R I Y P C L I L M P V F S F N I L K G M V R I G V V L A L S L M P A I G L C A
SSAT-SALTY ---M A Q Q V N E W L I A L A V A F I R P L S L S L L L F L L K S G S L G A A L L R N G V L M S L T F P I L P I I Y
YSCT-YEREN ---M I A D L I Q R P L L T Y T L L L P R F M A C F V I L P V L S K Q L L G G V L L R N G I V C S L A L Y V Y P A V A

```

```

FLIR-AGRO  T I Y P K V S G A S Q G A V V G L I F S E V L I G A M Y G L I A R E Y T L G F Q F T G A L I G A S I G L S A P G G A D P
FLIR-ECOLI M D V P V F S F ---F A L W L A W Q O T L I G I A L G F T M Q F A F A V R T A G E I I G L Q M G L S F A T F V D P
FLIR-BACSU  P P S L E I D E ---H Y M L L A F K E A L V G L C L G L I A Y M M I A V Q I A G S F I D F Q M G F S I A N V I D P
FLIR-ERWCA Q L A Y Q M P D W ---P Q L I G L I L K E V T I G I L L G L L L G M P F W L F Q S A G A L F D N Q R G A L I G G Q L N P
MOPE-ERWCA M T T P I F S V ---A G V W L L I Q O L L I G V T L G L S M Q L A F A A I R H A G E L I G L Q M G L A F A T F F D P
YOPT-YEREN Q L A Y Q M P D W ---P Q L I G L I L K E V T I G I L L G L L L G M P F W L F Q S A G A L F D N Q R G A L I G G Q L N P
SSAT-SALT  Q Q K I M M H I G K D Y S W L G L V T G E V I I G F S I G F C A A P F W A V D M A G F L L D T L R G A T M G T I F N S
YSCT-YEREN M O P Y I E V D A ---F T L M L L I G K E F I L G L L I G F V A T I P F W A L E S A G F I V D N Q R G A A M A S L L N P

```

```

FLIR-AGRO  I E D V Q E N Q I A N F I T F G G L L V L F M M D F H H I V L K A L V D S Y S A T P W G A ---L I S G Q K M L I T L T D
FLIR-ECOLI A S H L N M P V L A R I M D M L A L L L F L T F N G H L W L I S T L V D T F H T L E I G G ---E P L N S N A F L A L T K
FLIR-BACSU  Q T G A Q S P L I G Q F I Y T M A L L F M L S V N A H H L L I D G I Y Y S E Q Y I S W D Q A F E N F G D E Q F A Y F I A
FLIR-ERWCA A L G S D V T P L G L L I Q O T L I L L L I I G I G L S G I T Q I I W D S Y R I W P V L Q W L E L P H E E G F K O Y L A
MOPE-ERWCA T G G P N M Q V V A R F L N I L A I L L F L T F D G H L W M I S L L A D S E Y T L P L S ---Q S A P ---
YOPT-YEREN A L G S D V T P L G L L I Q O T L I L L L I I G I G L S G I T Q I I W D S Y R I W P V L Q W L E L P H E E G F K O Y L A
SSAT-SALTY T I E A E T S I F G L L F S Q F L C V I F F I S G G M E F I L N I L Y E S Y O Y L P P G R - T L I F D Q Q F L K Y I C A
YSCT-YEREN G L D S Q T S P T G L L L T Q T R I T I I F C G G A F L S L L S A L F H S Y V N W P V A S F F E E V S E Q W V D F F Y N

```

```

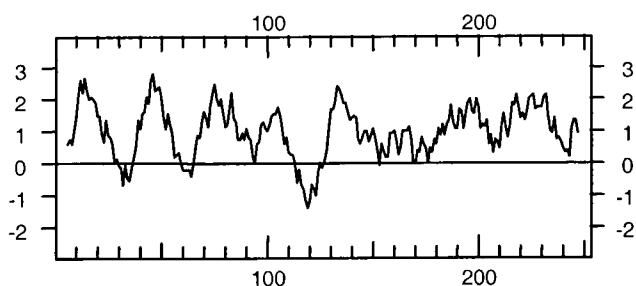
FLIR-AGRO  - T L R A S F S I M L R L A S P F V I Y G L M F N V A V G L I N K L A P Q I P V F F I S T P F V L A G G L F M L Y L S V
FLIR-ECOLI - A G S L I F I N G L M L A L P L T L L L T L N L A L G L L N R M A P Q L S I F W I G F P L T L T V G I S L M A A L M
FLIR-BACS  K S I N A M F I I A F Q M S A P V V A S L F L V D L A L G I V A R T V P Q L N V F V V G L P L K I A V S F I M L I V C M
FLIR-ERWCA - L L A D T E T H I I I Y A G P L V A L L L L D F S I A I L S L Y S P Q L Q V F V L S V P A K C L V G L L F F V L Y I
MOPE-ERWCA ---I L L T A M L F S L L L A P V D ---
YOPT-YEREN - L L A D T E T H I I I Y A G P L V A L L L L D F S I A I L S L Y S P Q L Q V F V L S V P A K C L V G L L F F V L Y I
SSAT-SALTY - E W R T L Y Q L C I S F S L P A I I C M Y L A D L A L G L L N R S A Q Q L N V F F F S M P L K S I L V L L T L L I S F
YSCT-YEREN - Q F S Q I L L I A A V L A A P L L I A M F L A E F G L A L I S R F A P S L N V F V L A M P I K S A I A S L L L V I Y C

```

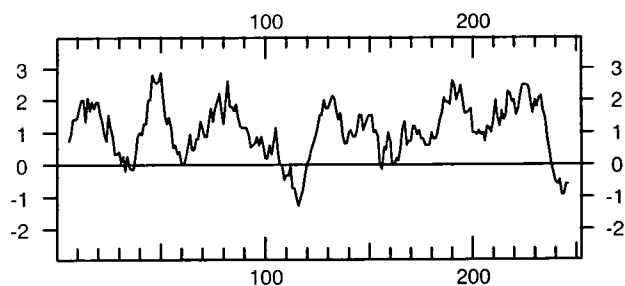
```

FLIR-AGRO  A G L I R Q F V D G F G P V F I G F Z ---
FLIR-ECOLI P L I A P F C E H L F S E I F M L L A D I I S E L P L I
FLIR-BACSU  S V I F V V V R N V F S L T I E T M R M L L A L V G V S
FLIR-ERWCA P T L N A L G E D R I L Q L R D L S K L L P L I L G G H
MOPE-ERWCA ---
YOPT-YEREN P T L N A L G E D R I L Q L R D L S K L L P L I L G G H
SSAT-SALTY P Y A L H H Y L V E S D K F Y I Y L K D W F P S V ---
YSCT-YEREN M Q M M S H A S K A M L L V M D P I S L L I P V L E K -

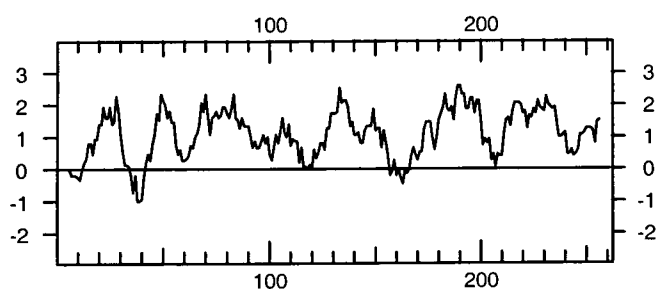
```



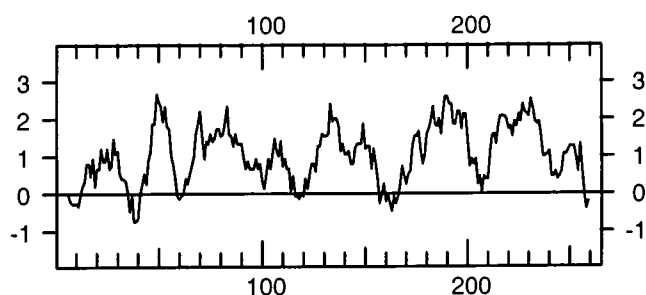
FliR from *A. tumefaciens*



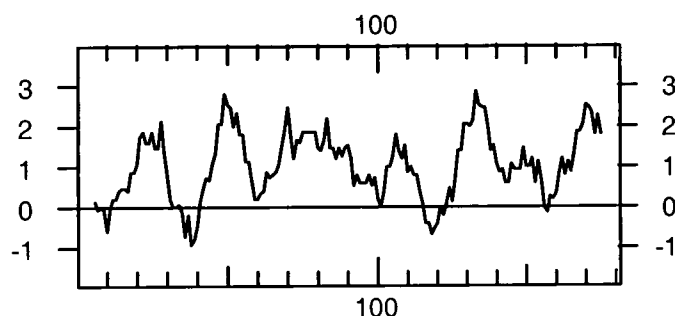
FliR from *E. coli*



FliR from *C. crescentus*



FliR from *S. typhimurium*



MopE from *Erwinia carotovora*

**Figur 3.6.3** Hydropathy profiles of the *A. tumefaciens* FliR homologue and FliR of *E. coli*, *C. crescentus*, *S. typhomurium* and MopE of *Erwinia carotovora* from type III export system .

*fliR* is a part of a large flagellar gene cluster containing seven additional genes encoding FliL, O, P, Q, R and B (75). The *Erwinia carotorova mopE* gene is in the *mop* operon (142). Mutations in this operon lead to a reduction of virulence in plants. In addition, a mutation in the *mopE* gene results in a non-flagellated phenotype. The *Shigella flexneri spaR* (*spa29*) gene resides in an operon (174), which encodes several genes that are essential for invasion of host cell and surface presentation of the virulence proteins. The function of *yscT* in *Yersinia pestis* has not been reported, however it is known that homologues of this gene in *Y. pseudotuberculosis* is essential for export of Yop proteins (92).

The hydropathy profile of the *A. tumefaciens* FliR homologue and the two other homologues, using the Kyte and Doolittle algorithm (shown in figure 3.6.3) revealed that they have similar patterns. It showed FliR to be very hydrophobic throughout almost the entire polypeptide sequence, and contain at least six membrane-spanning domains, suggesting that it is a membrane protein. The predicted transmembrane domains of the FliR, using SOSUI program at HGMP is shown in table 3.6.2.

### 3.7 *orf X*

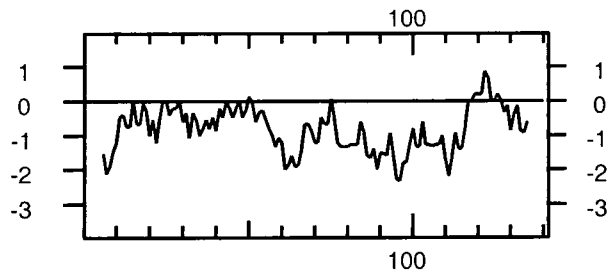
About 16 bp downstream of the *fliR* homologue an *orf* 756 bp long, starting at position 2982 bp and finishing at position 3401 bp was identified. The gene product of this *orf* (*orfX*) contains 139 amino acids, with molecular mass of 16.053 kDa. TESTCODE analysis predicted, at the 95% confidence level, that this *orf* is a coding region. The guanine and cytosine content of the *orf* is 54.7% and this is significantly higher than those of non-coding regions (see table 3.4.1). In addition, a potential ribosome-binding site was found 5 bp upstream of the start codon of *orfX*. However, the GenBank database search showed no significant homology to any previously identified proteins. This might be because no relevant protein has yet been found in other bacteria. The whole DNA sequence of *orfX* together with its possible protein sequence is shown in figure 3.7.1. The hydropathy diagram of ORFX, which has been shown in figure 3.7.2, has a very hydrophilic profile.

**Figure 3.7.1 The DNA sequence of *orfX* with its putative translated protein and DNA sequence upstream of *orfX*.**

*OrfX* starts at position 2983 bp and finishes at position 3401 bp. The initiation codon of the orf is underlined and labelled *orfX*, while the stop codons shown by "\*".

A potential ribosome binding site "GAGGA" 7 bp upstream of the start codon is underlined and marked "rbs".

	rbs					orfX																	
2964	tga	tcg	<u>gga gga</u>	aac	gcc	<u>ATG</u>	GCT	TCG	GAC	AAG	CGC	TCG	GAA	AAG	CTC	AAG	CGT	CTG	GTA	3023			
	*					<b>M</b>	A	S	D	K	R	S	E	K	L	K	R	L	V	14			
3024	ACG	GTC	CAG	CGG	CAT	ATG	GAA	AAA	ATG	GCC	GAG	GTG	GAA	CTG	GCC	GAT	ACC	ACT	CGC	GTG	3083		
15	T	V	Q	R	H	M	E	K	M	A	E	V	E	L	A	D	T	T	R	V	34		
3084	CGC	AGC	GAA	GTG	GCG	CAA	TCC	ATG	GAG	AGC	ACG	TTT	GAG	GCC	ATG	AGT	TCC	ATG	GAA	CCG	3143		
35	R	S	E	V	A	Q	S	M	E	S	T	F	E	A	M	S	S	M	E	P	54		
3144	GTG	CAT	CAG	ACC	TTC	TCC	AAG	CAT	TAT	TCG	GAC	CGC	TAC	AGC	CGC	CTG	GTC	GTC	CAG	GAC	3203		
55	V	H	Q	T	F	S	K	H	Y	S	D	R	Y	S	R	L	V	V	Q	D	74		
3204	CGC	CAG	CTC	GAA	GGC	GTT	CAG	CAG	TTT	CAG	GAA	AAC	AAG	GTT	CTG	AAG	GAA	AAG	ACC	AAG	3263		
75	R	Q	L	E	G	V	Q	Q	F	Q	E	N	K	V	L	K	E	K	T	K	94		
3264	GCT	GAC	AGG	CTG	GAG	GAC	AGG	ATG	CAT	ATT	GCC	CGC	GAT	CTC	GAG	GAT	CGC	GAG	GCG	GGT	3323		
95	A	D	R	L	E	D	R	M	H	I	A	R	D	L	E	D	R	E	A	G	104		
3324	GAT	AAT	GCA	ATC	TAC	GAT	TTG	TTA	GAA	ATT	ACG	AAT	GTT	TCC	CAC	ACA	CCA	GCC	TCC	AGC	3383		
115	D	N	A	I	Y	D	L	L	E	I	T	N	V	S	H	T	P	A	S	S	124		
3384	AAG	GTT	GGC	GAT	CCA	TAG															3401		
135	K	V	G	D	P	*															140		



**Figure 3.7.2** Hydropathy profiles of the ORF *A. tumefaciens*.

### 3.8 The *cheL* homologue:

The *cheL* homologue is a 532 bp open reading frame that in the TESTCODE analysis was shown to be a coding region. The *orf*, which spans the *che-2* mutant, starts at nucleotide 3644 (about 240 bp downstream of *orfX*) and finishes at nucleotide 4073 bp. The gene product of the *orf* is 144 amino acids long, with a predicted molecular weight of 15.66 kDa. It has sequence homology to the CheL protein of *C. crescentus* (28% identity and 53% similarity, over 75 amino acids). The *Caulobacter* CheL protein is the only CheL protein which has been found in the bacteria so far therefore this ORF is the second protein of this kind to be reported. The whole DNA sequence of *cheL* together with its possible protein sequence is shown in figure 3.8.1, while the GAP alignment of the possible *Agrobacterium* CheL, with the *Caulobacter* CheL, is shown in figure 3.8.2. Pairwise alignment between the two CheLs has revealed that there are four major conserved areas within the two proteins. The conserved regions are between residues 34 and 96 of the 111 amino acids. The hydropathy profile, according to the Kyte and Doolittle algorithm, of the two proteins is shown in figure 3.8.3. The CheL of *A. tumefaciens* is significantly longer than its counterpart in *C. crescentus* (about 60 amino acids).

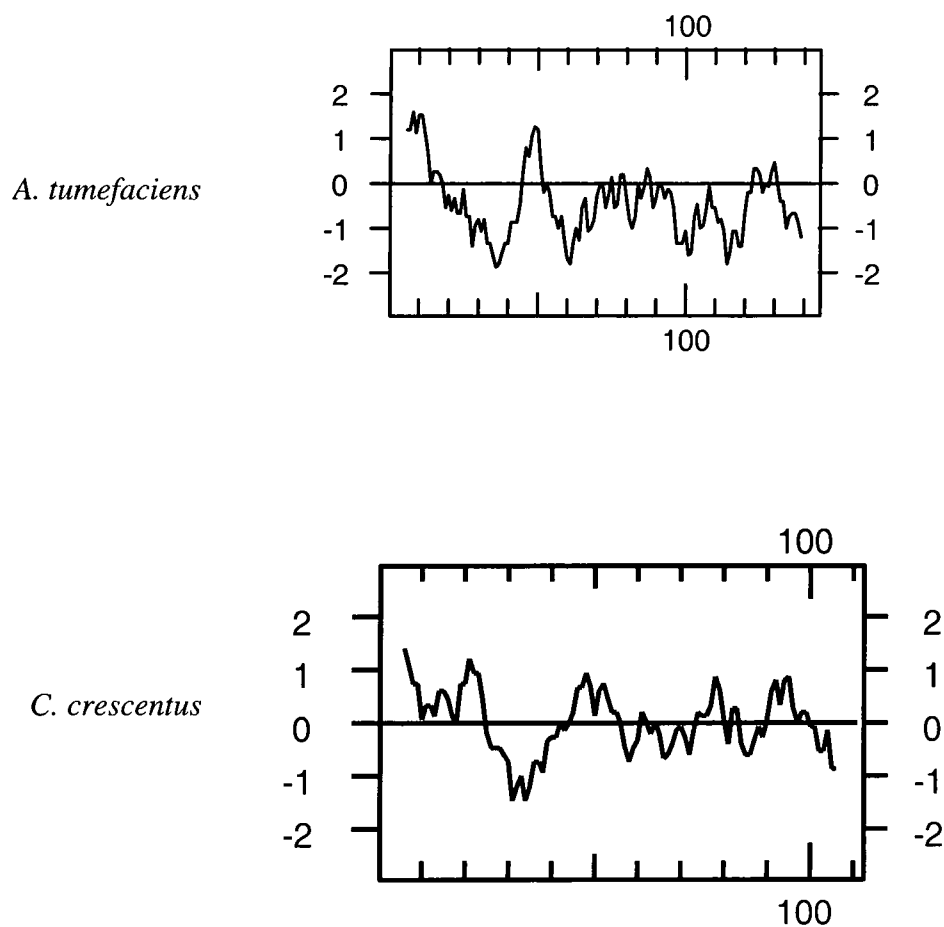
The DNA sequence upstream and downstream of *cheL* was analysed to try to find any promoter or terminator. No sequence indicating the presence of a terminator downstream of *orf* was found. A stretch of sequence, which overlaps the initiation codon of the *cheL* showed similarity to the *A. tumefaciens* putative flagellar promoter (GCCAATG) and -10 class II flagellar promoter from *E.coli* (GCCGATAA) (55), but no typical -35 consensus sequence preceded this sequence. A sequence (ACCAG) weakly resembling a Shine-Dalgarno motif was also found, 8 bp upstream of the start codon of *cheL*, this may therefore be a weak ribosome-binding site.







**Figure 3.8.3** Hydropathy profiles of the *A. tumefaciens* CheL homologue and its counterparts in *C. crescentus*.



**3.9. *orfY*, *orfW* and *orfZ*:**

The sequence downstream of *cheL* contained three open reading frames, *orfY*, *orfW* and *orfZ*. All open reading frames were predicted to be coding regions by TESTCODE analysis, at the 95% confidence level. Moreover they had guanine and cytosine content of 54.86%, 57.4% and 60.89% respectively, significantly higher than

that of non-coding regions (see table 3.4.1). However, the *orfs* have no significant sequence identity to any previously identified proteins in the GeneBank database.

*OrfY* begins just 2 bp downstream of *cheL* at position 4078 bp and finishes at position 4473 bp. It has 396 base pairs and contains 132 amino acids, with a predicted molecular mass of 14.72 kDa. A sequence (GAGG) found 7 bp upstream of the start codon could be a ribosome-binding site. No possible promoter or terminator was found upstream or downstream of *orfY*. The hydropathy profile shows that *orfY* is extremely hydrophilic.

The second open reading frame, *orfW* starts at position 4471 and ends at position 4923. This open reading frame contains 453 bp and 151 amino acids. The start codon of open reading frame *W* overlaps the stop codon of *orfY*, giving a "TGATG" configuration. Two motifs that resemble Shine-Dalgarno sequence are found 12 bp and 26 bp upstream of start codon of this *orf*; these may therefore be possible ribosome-binding sites.

There is a gap of 168 bp between *orfW* and the third open reading frame, *orfZ*, which was predicted as a non-coding region by TESTCODE analysis. A sequence-like-terminator, which consists of a 9 bp stem and a 3 bp loop was identified 47 bp upstream of *orfZ* in this region. If this sequence is a real terminator, it is therefore possible to postulate that *cheL* along with *orfY* and *orfW* constitute an operon as they overlap or are close to each other. Moreover, the existence of a possible -10 classII promoter-like sequence upstream of *cheL*, increases the probability of these being an operon in this region.

*OrfZ* starts at nucleotide 5191 and extends to the end of the sequence. It is 675 bp long and contains 225 amino acids. No typical ribosome-binding sequence was found upstream of the *orf*, however a TGGG that could be a ribosome-binding site, was identified 9 bp downstream of start codon. A -10 box classII promoter sequence, GGCGAT, with one bp mismatch and a -35 classII sequence, TCATGA, were also found 24 and 63 bp upstream of *orfZ*.

The DNA sequence of the three *orfs* and their deduced amino acids is shown in figure 3.9.1 to 3.9.3. The hydropathy profiles of these *orfs*, according to the Kyte and Doolittle algorithm, are shown in figure 3.9.4.

**Figure3.9.1 The DNA nucleotide and deduced amino acid sequence of *orfY* with untranslated region upstream of the *orfY***

*OrfY* begins at nucleotide 4077, two bp downstream of *cheL* homologue, and finishing at nucleotide 4472. Start codon is underlined and labelled "*orfY*", while stop codon indicated by asterisk. Stop codon of previous gene is also underlined and shown by double asterisks.

A potantial ribosome binding site is underlined and labelled "rbs"

4005	aat	tcg	gat	ggg	ggc	gga	ggt	cga	gcg	gca	ggt	cct	cgg	cgt	ttc	caa	tga	taa	aac	gga	4064
	rbs			orfY																	
4065	<u>cga</u>	<u>ggc</u>	<u>ttg</u>	<u>aga</u>	<u>ATG</u>	AAA	AAC	CTT	AAC	GAG	GAG	ACA	AGC	ATG	GAC	CTT	ATG	TCG	AAC	GAC	4124
1			**		M	K	N	L	N	E	E	T	S	M	D	L	M	S	N	D	16
4125	CAC	CGT	ATC	CAA	TCC	GTT	TTC	GGC	CGC	CTC	GAG	ATG	ATC	ATC	GAC	AAT	GAG	AAC	GAC	AAT	4184
17	H	R	I	Q	S	V	F	G	R	L	E	M	I	I	D	N	E	N	D	N	36
4185	ATC	GGT	AAG	GAC	CCG	CAG	TTC	GAC	TTC	AAG	GTC	TCC	AAT	GCC	CAT	AAG	AGC	CGT	TGC	CTT	4244
37	I	G	K	D	P	Q	F	D	F	K	V	S	N	A	H	K	S	R	C	L	56
4245	TAT	GAA	CTG	ACG	ATG	CTG	GTC	CGC	GAT	ACA	CCG	CAC	GAA	GAT	ATC	GCC	GCC	GGT	TAC	ATG	4304
57	Y	E	L	T	M	L	V	R	D	T	P	H	E	D	I	A	A	G	Y	M	76
4305	GAG	CAG	GTG	AAG	GGC	ATC	AAG	TCC	AAG	CTC	GCC	ACC	AAT	GCC	ACC	CGC	GTC	GAG	GCG	CAT	4364
77	E	Q	V	K	G	I	K	S	K	L	A	T	N	A	T	R	V	E	A	H	96
4365	CTG	AAC	GCG	GTT	CGC	GCT	GTC	GCG	GAT	CTG	CTC	AAG	AAC	GCT	ATT	CAG	GAA	GCC	GAT	ACG	4424
97	L	N	A	V	R	A	V	A	D	L	L	K	N	A	I	Q	E	A	D	T	106
4425	GAC	GGC	ACC	TAT	TCC	CAG	GAA	CAA	TTC	CTT	TAC	GGC	GCA	GCC	TCC	TGA					4472
107	D	G	T	Y	S	Q	E	Q	F	L	Y	G	A	A	S	*					122

**Figure 3.9.2 The DNA nucleotide and deduced amino acid sequence of *orfW* with untranslated region upstream of the *orf*.**

*OrfW* starts at position 4471, which overlaps the stop codon of *orfY*, and finishing at position 4472 bp. The start codon is underlined and labelled "*orfW*", while stop codon indicated by asterisk. Previous *orf* stop codon is underlined and shown by double asterisk. Two potential ribosome-binding sites 12 (ACGG) and 26 (AGGA) bp upstream of the *orf* start codon are underlined and labelled "rbs".

													rbs														rbs			
4400	gaa	cgc	tat	tca	gga	agc	cga	tac	gga	cgg	cac	cta	ttc	cca	<u>a gga</u>	aca	att	cct	tta	<u>cgg</u>	4459									
<i>orfW</i>																														
4460	cgc	agc	ctc	<u>ctg</u>	<u>ATG</u>	TTG	AAG	CTT	CTT	CTC	ACC	GGC	GTC	TGG	GTT	TGC	GCC	GTC	ACG	CTC	4519									
1				**	<b>M</b>	L	K	L	L	L	T	G	V	W	V	C	A	V	T	L	16									
4520	GGC	GCG	GTG	TAT	TTC	TCC	GTC	CAG	ATG	GCG	ACG	GCG	CCG	GCG	CCG	GAT	GAG	GCG	GGG	GCA	4579									
17	G	A	V	Y	F	S	V	Q	M	A	T	A	P	A	P	D	E	A	G	A	36									
4580	AAA	AAG	GCT	GAT	CTG	CAA	CTG	GTC	AAG	GGC	GAA	AGC	ATC	ACG	ATC	CCC	GTC	ATC	AAT	GAT	4639									
37	K	K	A	D	L	Q	L	V	K	G	E	S	I	T	I	P	V	I	N	D	56									
4640	GGT	GGA	GTG	AAC	GGC	TAT	TTC	CTC	AGC	CGG	ATT	TCG	CTG	CGC	GTC	GAC	AAG	GCC	AAG	ATG	4699									
57	G	G	V	N	G	Y	F	L	S	R	I	S	L	R	V	D	K	A	K	M	76									
4700	GCG	AAG	ATC	GAA	CTG	CCG	GCA	ACA	CAG	CTG	ATG	ACC	GAT	GAA	CTG	TTC	ACG	CTT	CTG	GCG	4759									
77	A	K	I	E	L	P	A	T	Q	L	M	T	D	E	L	F	T	L	L	A	96									
4760	GGT	TCT	TCC	ATG	GTC	AAC	ATT	GCC	AAT	ATC	TCC	ACC	TTC	GAT	CCC	GAA	GCC	TTC	AAG	CAG	4819									
97	G	S	S	M	V	N	I	A	N	I	S	T	F	D	P	E	A	F	K	Q	116									
4820	CGC	ATC	CGC	GAA	GGG	TTG	AAC	AAG	AGG	CTC	GAT	GAC	GAG	GTC	GTC	GAG	GAC	GTG	CTG	ATC	4879									
117	R	I	R	E	G	L	N	K	R	L	D	D	E	V	V	E	D	V	L	I	136									
4880	GAG	CAG	CTC	GAT	TAC	CTG	TTC	GAA	AGC	CGA	CAT	CCG	CGA	GCA	GAA	GGA	CAA	TGG	TTC	GCC	4939									
137	E	Q	L	D	Y	L	F	E	S	R	H	P	R	A	E	G	Q	W	F	A	156									
4940	CCA	CTC	TGT	CAA	ACT	CGT	CGA	GGG	CGA	AAA	AGC	CGA	AGC	AGG	AGA	GGC	TGC	GGC	ACC	CAG	4999									
157	P	L	C	Q	T	R	R	G	R	K	S	R	S	R	R	G	C	G	T	Q	176									
5000	CCA	CTG	ACG	GAG	AGG	GAG	CGG	CGG	CGA	AAC	GGC	ACG	CTG	GTT	AAT	CAT	CTC	TTG	CTC	CAT	5059									
177	P	L	T	E	R	E	R	R	R	N	G	T	L	V	N	H	L	L	L	H	196									
5060	TCA	CGA	GCA	TTC	GAG	AGA	TTG	CTT	TCA	AGT	TAA															5093				
197	S	R	A	F	E	R	L	L	S	S	*															208				

**Figure3.9.3 The DNA nucleotide and deduced amino acid sequence of *orfZ* with untranslated region upstream of the *orf*.**

*OrfZ* starts at position 5188 and finishing at position 5731 bp. The start codon is underlined and labelled "*orfZ*", while stop codon indicated by asterisk.

A potential ribosome binding sites (TGGG), 8 bp upstream of the orf start codon are underlined and labelled "rbs". Approximately 24 and 63 bp upstream of start codon are a putative –10 and –35 promoter sequences underlined and labelled "-10 and –35".

The potential stem-loop structure, which may be a transcription termination signal, is also shown between nucleotides 5140-5160 by staggered arrows and labelled "TM".

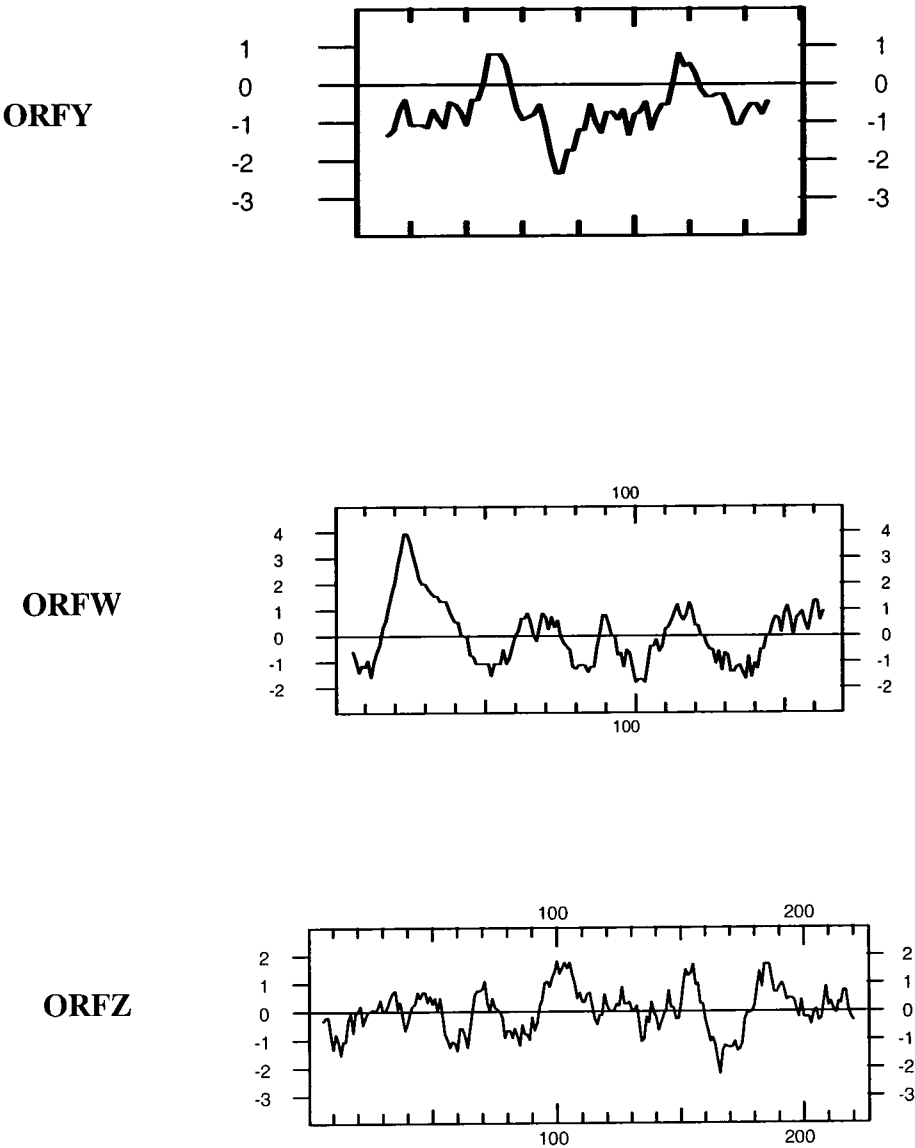
		-35																			
5092	aat	tca	acc	tgc	acc	tgt	ata	aca	aac	<u>tca tga</u>	acc	acc	gcc	ccg	cag	gcc	atc	CGC	CGG	5152	
										----- -->TM											
		-10				rbs				orfZ											
5152	gcc	ggt	<u>ggc gat</u>	cgg	ttt	ggc	att	<u>tgg ggg</u>	tgc	gca	<u>ATG</u>	AAT	TTC	GCG	TCA	GGT	GTT	ACG	5211		
1	←	-----									M	N	F	A	S	G	V	T	8		
5212	CGC	CGG	GGA	TTG	CAA	CGG	AAT	ATC	CAC	GGG	CTG	CGC	GTC	TGT	GAT	CTC	GAC	TGG	AAT	GGC	5271
9	R	R	G	L	Q	R	N	I	H	G	L	R	V	C	D	L	D	W	N	G	28
5272	GCG	CTG	GAA	CTG	GTC	AGC	GGC	AAG	GCT	TCG	GCC	TGT	GAA	GGG	CAC	ACG	ATG	CTC	TCC	TTC	5331
29	A	L	E	L	V	S	G	K	A	S	A	C	E	G	H	T	M	L	S	F	48
5332	CTG	ACC	ATC	GAG	AAG	ACA	AAG	TTG	TCA	CTC	AAG	GAT	AAT	GCC	TAT	CGG	CAA	ATT	CTG	GAA	5391
49	L	T	I	E	K	T	K	L	S	L	K	D	N	A	Y	R	Q	I	L	E	68
5392	AGC	TGC	CTG	CTT	CTG	CCG	CAG	GGC	AAG	GGC	ATG	AAA	GCC	GCC	GCG	CGG	GCC	GAA	CAT	GGC	5451
69	S	C	L	L	L	P	Q	G	K	G	M	K	A	A	A	R	A	E	H	G	88
5452	GAG	CCT	CTG	CCC	GCC	ACC	ATC	GAT	GGC	GTT	GCC	TTC	ACC	ATG	GCG	CTT	CTG	ACC	TAC	ATG	5511
79	E	P	L	P	A	T	I	D	G	V	A	F	T	M	A	L	L	T	Y	M	108
5512	GCT	GTG	CCG	AAA	CGC	ATC	GGC	GTG	GCG	GGC	GAC	GAT	GCG	GCG	CAG	ATC	GGT	GAA	GTG	CTT	5571
99	A	V	P	K	R	I	G	V	A	G	D	D	A	A	Q	I	G	E	V	L	128
5572	GCC	CGG	CTG	CGT	GCC	CAT	GCG	CCA	TGG	CAA	GAC	TTC	GTC	ATG	CTG	AAC	CGG	GAC	GCG	GCC	5631
119	A	R	L	R	A	H	A	P	W	Q	D	F	V	M	L	N	R	D	A	A	148
5632	GGC	GTG	AAG	GTG	GAT	GTG	CTC	CTG	GCA	GGA	ATG	CTC	AAG	GAA	AAT	CAG	GAG	ACA	TGG	CTG	5692
139	G	V	K	V	D	V	L	L	A	G	M	L	K	E	N	Q	E	T	W	L	168
5692	CAC	CGC	AGC	GTC	AGC	CGT	GAG	GAT	GCG	CGT	GTC	ACC	ATT	GCA	GTC	GGG	CCC	CTG	TTC	AAG	5731
159	H	R	S	V	S	R	E	D	A	R	V	T	I	A	V	G	P	L	F	K	188
5752	GTC	CTG	TCC	TCA	GAA	GTG	GCG	GGT	ATG	CCG	GAC	ATT	TTC	CGC	AAG	CTG	CAT	ATG	AGC	TGG	5811
179	V	L	S	S	E	V	A	G	M	P	D	I	F	R	K	L	H	M	S	W	208

5812 CTC TAC AGC CTT TGT GCC GAA CCC TGG CAC ATC GCG CTC GGC AAG AGC TGA

199 L Y S L C A E P W H I A L G K S \*

5865  
225

**Figure3.9.4** Hydropathy profiles of the three unknown ORFs , ORFY, ORFW and ORFZ in *A. tumefaciens*.





### 3.10 Discussion

Previous work in this lab showed that the chemotaxis and motility genes are clustered together in *A. tumefaciens*. They showed three cosmids, pDUB1900, pDUB1905 and pDUB1911 to contain many of the behavioural genes(34, 180, 228). The genes on the cosmids pDUB1900 and pDUB1911 have already been reported (56, 57, 228). The data presented here describes the identification and sequencing of two flagellar homologues, one chemotactic gene and four open reading frames encoding potential proteins without significant sequence similarity to any other proteins in the GenBank database. All are located on the cosmid pDUB1905.

Based on data obtained from sequence analysis, it is possible to postulate that the flagellar and chemotaxis homologues along with the three unknown open reading frames, *orfX*, *orfY* and *orfW* are transcribed together. This hypothesis is supported by the existence of a transcription termination sequence at the start of the sequence and another terminator in the end of the *orfW* sequence. However no classical promoter was found upstream of the possible operon. It is also tempting to speculate that the sequence can be divided into at least two distinct operons. One operon consists of the *cheL* homologue and the two unidentified *orfs*, *orfY* and *orfW*. The *flhA* and *fliR* genes with or without *orfX* could constitute a second operon, or may be members of a larger operon. This speculation comes from the fact that a rather big gap exists between the two assumed operons and the genes in the operons overlap or are close to each other. The presence of a terminator upstream of *flhA*, a -10 element class II promoter-like sequence upstream of *cheL* and a terminator downstream of *orfW* are further indications of the existence of these operons. On the other hand, no promoter was found upstream of *flhA*, this might be because the sequence upstream was not enough to reveal any promoter, or relevant genes. In *C. crescentus*, it has been shown that a sigma 54 promoter controls transcription of *flhA* (171). Also no terminator was found downstream of *fliR* or *orfX*. The last *orf*, *orfZ*, appears to belong to another operon or is translated individually, since a possible terminator and a promoter was found upstream of this open reading frame. Sequencing downstream of this *orf* and upstream of *flhA* would be valuable to try to determine whether or not *orfZ* or *flhA* are part of an operon, by looking for possible genes or other information. It has been shown that the *Agrobacterium* motility and chemotaxis gene cluster has considerable

similarity with the closely related bacterium from the  $\alpha$ -subgroup, *S. meliloti* (190). Most of the flagellar genes (*fla*, *flg*, *flh*, and *fli*) and motility genes are in the same order in the two clusters. However, there is a difference between the two clusters. The *che* operon in the *Sinorhizobium* cluster is linked downstream of *flhB*, while downstream of the *Agrobacterium flhB* are the fourth flagellin gene, *flaD* and the *motBCD* operon. Despite this difference between the two clusters, the gene order within each group (*fla*, *fli*, *flh* and *che*) appears identical. Therefore, it is possible to predict that the order of genes found in this work is also the same in both bacteria. Although, *flhA*, *fliR* and *cheL* have not yet been identified in *Sinorhizobium*, but it has been shown that a fragment from *Agrobacterium* containing the *che-2* mutant, hybridised to a 3.7 kb fragment of *S. leguminosarium* biovar *viciae* (55).

The protein product of the *A. tumefaciens flhA* homologue had significant sequence similarity to the FlhA proteins of several bacteria in GenBank (see section 3.5). The predicted *flhA* product has similar features with other FlhA proteins. It is slightly acidic (pI=6.679), and has 703 amino acids that is in the range of other FlhA proteins (Table 3.5.). The amino acid sequence is divided into two domains with the N-terminal half consisting of six strongly hydrophobic regions that are predicted to be a membrane-spanning segment, it is therefore predicted to be an integral membrane protein. On the other hand, the carboxy-terminus end of FlhA is hydrophilic. The high similarity between *Agrobacterium* FlhA and other FlhAs was found in the N-terminal half. *A. tumefaciens* is therefore predicted to have the same structure and function as FlhA in other bacteria. *flhA* is classified as a class II gene of the regulatory hierarchy genes in *E. coli* and *S. typhimurium* (122). This class of genes is required for flagellar assembly and function. FlhA along with at least eight other genes constitute the flagellum-specific export apparatus (see section 1.4.3), which are responsible for transport of the protein components of the extracellular structures, such as the hook and filament proteins, across the membrane. It has been shown that FlhA in *S. typhimurium* is a membrane-associated protein. There is evidence that FlhA is associated with the basal body. FlhA has large soluble domains that presumably protrude into the cytoplasm within the cavity that lies at the centre of the cytoplasmic C ring of the motor (138). It has been demonstrated that FlhA has a regulatory activity in addition to its role in flagellar transportation. In a nonmotile mutant of *Proteus* lacking FlhA, expression of classII and classIII flagellum genes is suppressed (80).

The protein product of the *A. tumefaciens flhA* homologue also showed significant homology to a family of typeIII proteins involved in virulence from a variety of animal and plant pathogens (see table 3.5.2). The similarity between proteins involved in the flagellum-specific export pathway in several gram-negative and gram-positive bacteria, and proteins from typeIII secretion systems is not surprising given, they share similar overall structural features and are involved in transport and localisation of the component proteins of the extracellular structure across the cell membrane by a mechanism that does not require signal cleavage. They have high similarity in their hydrophobic amino-terminus domain and less similarity in their C-terminus domain. Perhaps the carboxy-terminal of these proteins functions in different manners in the diverse environment that the bacteria live. InvA from *S. typhimurium* is required for the ability of *Salmonella* to invade cultured epithelial cells. Secondary-structure analysis of InvA (74) predicts that there are seven transmembrane domains located in the amino-terminal half of the protein, and a largely hydrophilic carboxy-terminus that is located in the cytoplasm (74). LcrD, which exhibits a virulence property called  $\text{low-Ca}^{2+}$  response (lcr), is necessary for expression of genes encoding a number of Yop proteins that are required for virulence and host invasion. It has been shown that LcrD contains eight possible membrane-spanning segments in its hydrophobic amino-terminal region (159). Other proteins such as MxiA of *S. flexneri* (former VirH), which is necessary for secretion of the Ipa invasion proteins, and HrpO, HrpI and HrpC2 that are all involved in virulence, have the same structure as LcrD (92). Although, members of this family appear to have similar structures, they are involved in diverse cellular processes, including virulence in *Y. pestis*, host invasion in *S. typhimurium*, and flagellar biogenesis in *C. crescentus*, *E. coli* and *B. subtilis*. Taken together, the size, the sequence homology, and protein structural features between the possible *Agrobacterium* FlhA, other FlhA and virulence proteins in the other bacteria, leads to the conclusion that this region of pDUB1905 contains the *A. tumefaciens flhA* gene.

The predicted protein product of *Agrobacterium fliR*, which is located immediately after *flhA*, showed similar features to FliRs of several other bacteria (in addition to its strong homology to FliRs). FliR of *E. coli* is a 261 amino acid protein and is extremely hydrophobic, containing at least five membrane-spanning domains. The possible *Agrobacterium* FliR also contains 261 amino acids that are within the range of FliRs (251 aa *C. crescentus* and 269 aa in *R. sphaeroides*). The hydropathy

profile of the predicted FliR is similar to those of other FliRs. It is very hydrophobic throughout almost the entire polypeptide sequence. Hydropathy analysis revealed the *Agrobacterium* FliR to have at least 5 membrane-spanning domains. The similarities between *A. tumefaciens* and other homologues suggest that this polypeptide likely folds into a similar protein structure and serves similar functions. It has been suggested that the location of the N-terminal domain of the FliR protein is probably cytoplasmic. This protein is known as a class II flagellar gene residing near the top of the regulatory hierarchy that determines the order of flagellar gene transcription (122). It has been shown that FliR is a membrane protein, involved in early step of flagellar biogenesis and like FlhA, belongs to a family of proteins that constitute the flagellum-specific export pathway.

The predicted *Agrobacterium* FliR protein like FlhA also has homology with a family of proteins (typeIII secretion proteins) that are involved in transport of virulence factors (see table 3.6.1). This is again not surprising, because the members of both families are involved in the same activities of exporting the flagellar or virulence proteins that do not contain cleavable peptide signals. For example, SpaR (Spa29), which is essential for invasion of host cells and surface presentation of the virulence factor, or proteins like Ipa, harpin, Yops, InvE, and some flagellum that are exported by these two specific pathways do not contain the conserved N-terminal signal sequence. It seems that this superfamily of proteins is used for export or assembly, or both, of proteins that lie beyond the cell membrane.

Many components of the flagellum-specific export pathway have already been identified in *A. tumefaciens* in this laboratory (*fliI*, *fliP*, *flhB* and *fliN*), and are located in cosmid pDUB1900 (56). Now, two more possible members, *flhA* and *fliR*, of this system are being reported, located on a different fragment of the chromosome, which is located close to the cosmid pDUB1900. More components of the export pathway in *A. tumefaciens* are still required to be identified. *fliO*, *fliH* and *fliQ* have not yet been found in *Agrobacterium*. Finding these members would give a full prospective of function of the flagellum-specific export pathway and relation between the members.

Many bacteria have both protein delivery systems for pathogenic factors and flagella units, however this is not the case for *Agrobacterium*. It is that *A. tumefaciens* does not possess a typeIII secretion protein system in despite finding such system in *S. meliloti*. Since the mechanism of pathogenicity that is used by *Agrobacterium* is different from those bacteria which possess typeIII secretion pathways involved in

delivery of virulence proteins. *Agrobacterium* pathogenicity is brought about by insertion of a piece of plasmid DNA (T-DNA) instead of virulence factors. This piece of DNA contains the genes required for producing tumours and opines. Transformation of plasmid T-DNA from *Agrobacterium* into host cells, mediated by a mechanism that is referred to as type IV system.

The protein product of the sequence that spans the *che-2* mutant in *A. tumefaciens* showed similarity to CheL of *C. crescentus*. However the size of *A. tumefaciens* CheL is longer than that of *C. crescentus* (101). Unfortunately little is known about CheL and hence its precise function is unclear. It has been shown that *cheL* can complement a mutant possessing a functional flagellum, but unable to respond to chemotactic signals. The same phenotype was seen in *Agrobacterium che-2* mutant. Since the characteristics of this mutant are similar to those of the *che-2* mutant, it is possible to conclude that the two genes have similar functions, in addition to their homology in their structures. The middle part of the two proteins was found to have higher similarity; this may indicate the importance of the central region of the proteins for function. It has already been shown that at least the last 5 codon of *cheL* are not critical for function, as deletion of these does not affect the ability of the gene to complement a *cheL* mutant (101). The size difference between the two *cheL*s may reflect the evolutionary divergence between two bacteria. Another possibility is that this *orf* in *A. tumefaciens* might not be a *cheL* gene, instead is a unique gene, specific to *A. tumefaciens* with *cheL* nature and structure.

Unlike the flagellar and chemotaxis gene homologues, it is impossible to infer functional roles for the gene products of *orfX*, *orfY*, *orfW* and *orfZ* as a result of homologies to previously identified proteins. This may be because the gene homologues to these *orfs* have not yet been identified in any other species. Two open reading frames encoding proteins with unknown functions and no homologies to previously identified proteins have also been reported within a flagellar operon of *B. subtilis* (36). Such an open reading frame, *orfB*, has also found in a cluster of flagellar genes in *A. tumefaciens* (57). There is a possibility that these *orfs* encode proteins involved in functions or structural roles specific for *A. tumefaciens*, which are not found in other bacteria. However, their position within a region that contains two components of the flagellum-specific export pathway suggests that they may also be involved in this pathway. Considering the characteristics of *orfs*, *X*, *Y* and *W*, it is tempting to postulate that they may have chaperone functions (see section 1.4.3). All

of three ORFs share some of chaperone properties. They have molecular weight in the range of typeIII flagellum export proteins and typeIII secretion system (15-20 kDa), acidic isoelectric point (pI, see table 3.4.1). Since the activity of chaperones is specific to one or a few substrates, they do not usually show sequence similarity to other sequences. This could put more weight on speculation of chaperone-like activity of three ORFs, as they did not show to have homology to any other genes in database. On the other hand, they surround the only chemotactic gene in this region, as can also be seen in *C. crescentus* where other flagellar genes surround *invA* gene. Similarities in size, molecular mass, isoelectric point and guanine and cytosine content of all the *orfs*, except *orfZ*, with *cheL* should be considered as another possibility that they have a *cheL*-like nature and are possibly chemotactic genes. It is interesting to note that CheL has also the same chaperone properties as the ORFs (see table 3.4.1). This suggests that CheL might have chaperone-like activity as a flagellar or chemotactic protein too. This is unlikely, since evidence suggest that *cheL* is more likely to be a chemotactic gene rather a flagellar gene, and no chaperone has yet been reported for a chemotaxis system. Chaperones are specific to the delivery of proteins from inside the cell, across the membrane to the surface or beyond the cell, while the chemotaxis system transfer the extracellular signals from outside to inside the cell through a series of protein-protein interactions. It is therefore, unlikely to find a chaperone for chemotaxis system. More experiments should be carried out to determine the role of the *orfs* and *cheL* identified in this work. In the first place, mutation analysis experiments should be conducted, followed by phenotypic observation to find out whether this has any effect on motility or chemotaxis. Protein expression studies are also required to show that the gene products of the *orfs* expressed. Any information obtained from the overexpression of the *orf* gene products could also reveal the roles of these putative proteins. If the proteins are expressed, then immunocytochemistry could be carried out to localise their position in the cell and, thus, helping to find their possible functions.

The results obtained in this part of the project identified more behavioural genes in *A. tumefaciens* that are located on fragment of chromosomal DNA which is different from the fragments containing other motility and chemotaxis genes previously found in this laboratory. Since no flagellar or chemotaxis gene was found downstream of *cheL* and the fragments that contain the other behavioural genes (pDUB1900, pDUB1911 and pDUB1905) do not overlap each other, it may be useful

to work on the proximity of these by finding cosmids from an *A. tumefaciens* library that overlap them. Resulting information can then explain the relation between the genes, whether they are clustered in one large fragment or there are intergenic gaps between the clusters. It would also give the location of the cosmids on the chromosome and would determine the exact distance between their *Bam*HI sites.

## **4. Mutagenesis**



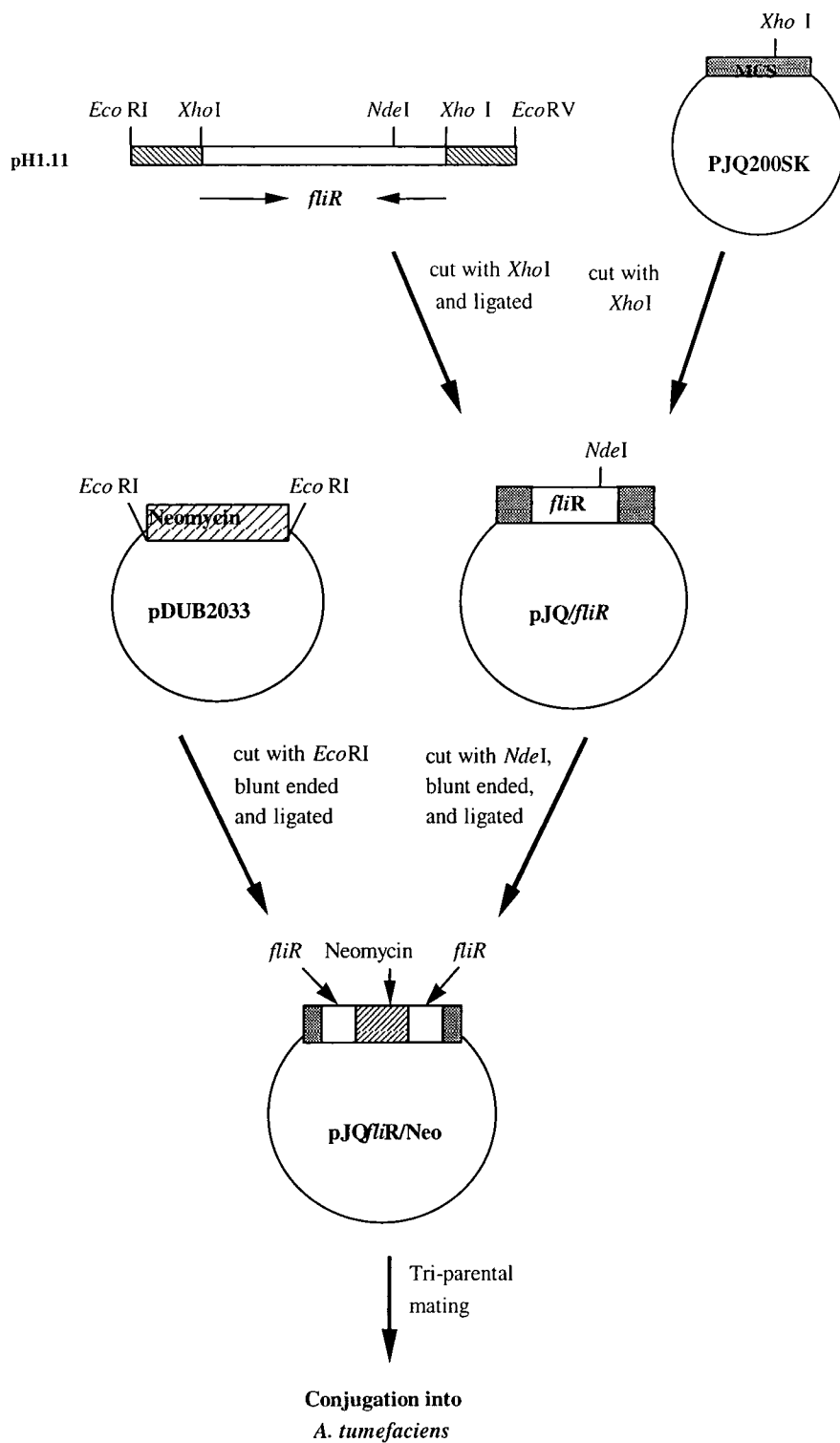
#### 4. Mutagenesis of *fliR*:

In an attempt to determine the effect on motility of one of the flagellum-specific export proteins of *A. tumefaciens*, strain C58C<sup>1</sup>, *fliR* was mutated by the gene replacement mutagenesis method using a neomycin cassette. The gene in question was knocked out by insertion of a neomycin cassette within the gene. The effect of the disrupted gene on motility was analysed phenotypically under light and electron microscopy and by swarm plate analysis. The gene replacement mutagenesis method was performed according to the protocol of J. Quandt and M. F. Hynes (162). The experimental procedure and basic theory are explained in section 2.10. A diagram of the procedure used to create the *fliR* mutation is shown in figure 4.1.1 This experiment involved two main phases. The first phase was to create a plasmid containing the gene disrupted by insertion of the neomycin cassette. The second phase involved introduction of the plasmid into *Agrobacterium*.

##### 4.1 Creation of the *fliR* mutant plasmid:

Information obtained from sequencing showed that *fliR* lies between two *XhoI* sites of the plasmid ph1.11. The first step therefore was isolation of *fliR* from ph1.11. This plasmid was digested with *XhoI* and a 0.8 kb fragment was cut out. The 0.8 kb fragment contains the complete *fliR* sequence plus a few bases of SK<sup>+</sup>. The next step was to clone the *fliR* fragment into the suicide vector pJQ200SK (called pJQSK for simplicity). The vector was cut with *XhoI*, purified and the 0.8 kb fragment (ph1.11/*XhoI*) was then ligated into the vector, forming the plasmid pJQ*fliR*. Positive colonies containing the *fliR* fragment were selected based on inactivation of the  $\beta$ -galactosidase gene present within pJQSK, and confirmed by subsequent restriction enzyme digests of the possible correct colonies. Plasmid pJQ*fliR* contains a unique *NdeI* site, located approximately in the middle of the second half of the *fliR* fragment (see figure 4.1.1). The plasmid was digested with *NdeI* and bluntended.

The third step was disruption of *fliR* gene by inserting the neomycin resistance cassette within it. The 1.2 kb cassette was obtained from the plasmid pDUB2033 (see table 2.2.3.2) by digesting with *EcoRI*. The cassette was blunt ended and then



**Figure 4.1.1** Diagram outlining the main sub-cloning steps involved in the construction of the *fliR* mutant by insertion of neomycin cassette. (see text for details)

The empty box represents *fliR*, while the light shaded box indicates Neomycin cassette.

subcloned into blunt ended *Nde*I cut, pJQ*fliR*. Since pJQSK contains a gentamycin resistance gene, correct colonies were selected for by growth on LM-agar plates containing both gentamycin and Neomycin, and confirmed by subsequent restriction enzyme digests. The plasmid was named pJQ*fliR*/Neo.

#### 4.2 Construction of an *A. tumefaciens fliR* mutant strain:

The second phase of the mutagenesis was to conjugate the plasmid pJQ*fliR*/Neo into *A. tumefaciens* C58C<sup>1</sup> by tri-parental mating (see section 2.9). Nine colonies, which grew on neomycin, rifampicin and sucrose, were selected and subcultured in RifNeo5% sucrose plates.

Two Southern blots (see section 2.7) were carried out to find out whether the nine possible mutant colonies contained the neomycin cassette. Genomic DNA of the possible mutants was extracted and digested with *Hind*III. Wild-type *Agrobacterium* C58C<sup>1</sup> was also cut with the same enzyme. The neomycin cassette and the *fliR* fragment were radiolabelled and used as probes. The neomycin probe was expected to hybridise only with the mutated DNA, while the *fliR* probe was expected to hybridise to both wild type and mutant DNA, but to different sized bands. The Southern blot hybridisation results can be seen in figure 4.2.1. When the blot was hybridised with the neomycin probe (see figure 4.2.1a), there was no band for wild-type C58C<sup>1</sup>, and two bands, of the expected size for the two mutant colony lanes. The *fliR* probe (see figure 4.2.1.b) also gave the predicted bands, and hybridised to both wild type and mutant DNA (see figure 4.2.2).

One mutant in the lane 3 appeared to have an extra band with the same size of wild type meaning that the mutant probably is not pure. The mutant phenotype was analysed in swarm plates and the result showed that the swarming pattern of mutant is as the wild type (5 cm after three days) indicating that this colony is a mixture of wild type and mutant. This mutant was discarded.

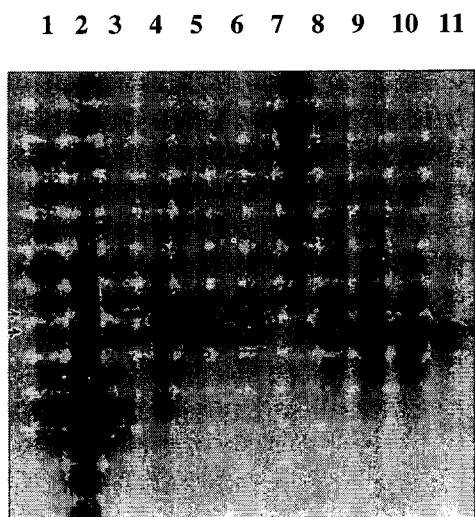
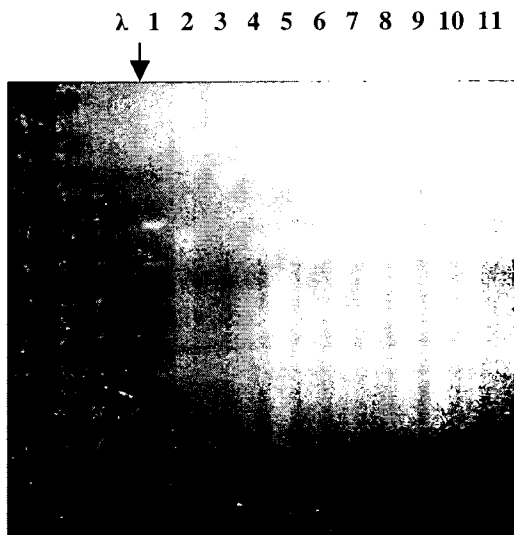
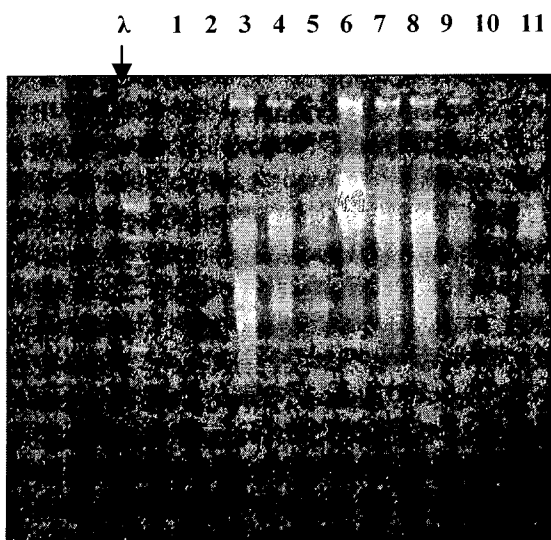
Because the Southern blot gave the expected results, one of the remaining colonies was selected for further analysis and renamed *fliR*Neo.

**Figure 4.2.1 Agarose gel of the genomic DNA cut with *Hind*III and subsequent Southern blot of nine putative *fliR* Agrobacteria mutant strain hybridised with *fliR* fragment (a) and neomycin cassette (b).**

(a) Lane 1 is the wild type *A. tumefaciense* (C58C1) genomic DNA digested with *Hind*III; lane 2 is *fliR* fragment used as control; and lane 3 to 11 are mutants genomic DNA cut with the same restriction enzyme.

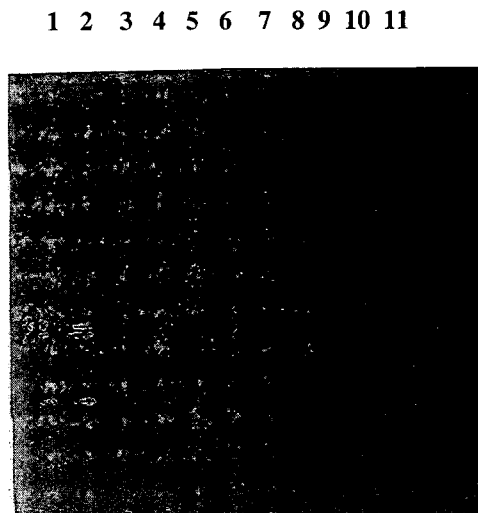
(b) Lane 1 is the wild type *A. tumefaciense* (C58C<sup>1</sup>) genomic DNA cut with *Hind*III; lane 2 is neomycin cassette fragment used as a control; and lanes 3 to 11 are mutants genomic DNA cut with the same restriction enzyme.

$\lambda$  represents  $\lambda$ -*Hind*III marker.



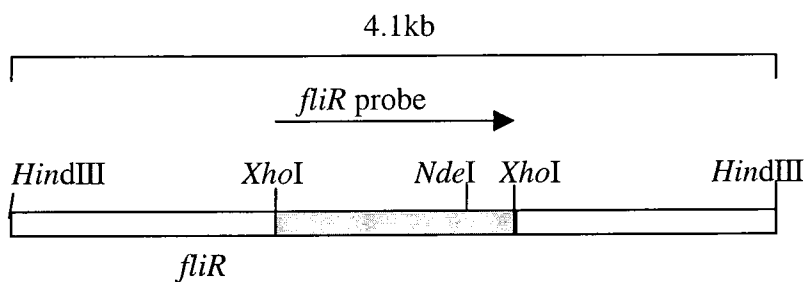
← 4.1  
 ← 2.8 →  
 ← 2.5 →  
 ← 1.1 →  
 ← 0.8

(a)

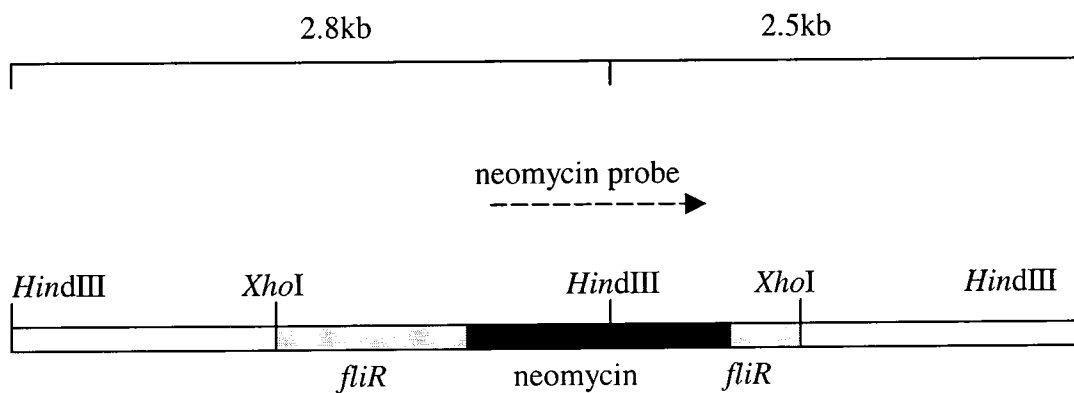


(b)

a):



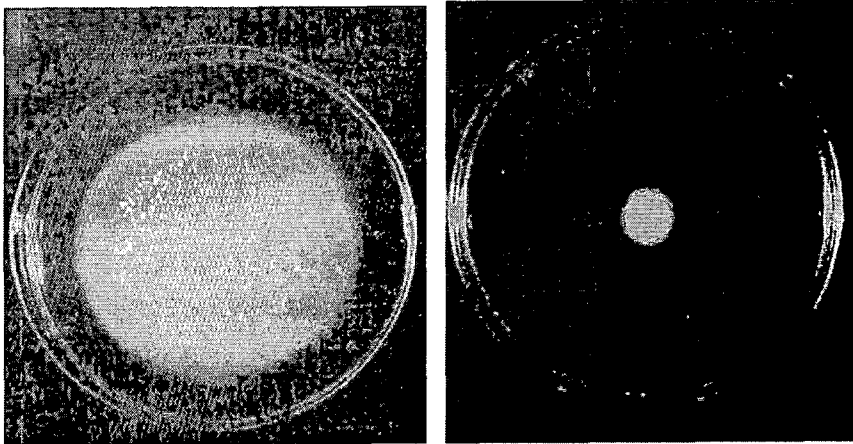
b):



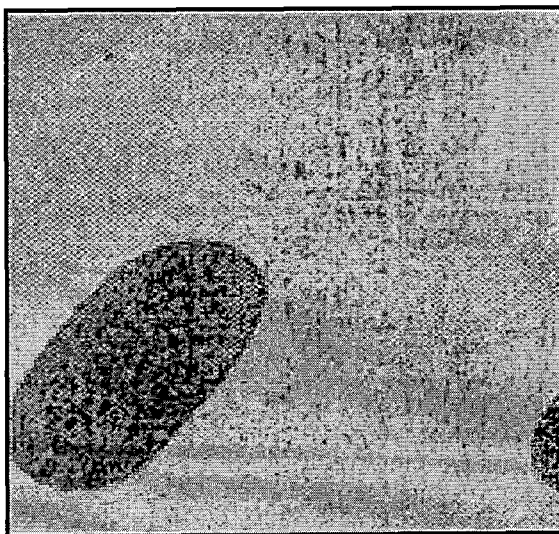
**Figure 4.2.2** Diagram showing the position and approximate size of the hybridising fragments of *fliR* and neomycin cassette to C58C<sup>1</sup> and *fliR* mutant strain.

a): Hybridising *fliR* probe with C58C<sup>1</sup> wild type produces a 4.1 kb fragment.

b): Hybridising *fliR* or neomycin cassette to *fliR* mutant produces two bands of 2.8 and 2.5 kb due to *Hind*III site in neomycin cassette.



**Figure 4.3.2** Photograph showing the swarming behaviour of the *fliR* mutant (right) and C58C1 (left) as a control.



**Figure 4.3.3** Electron micrograph of the *fliR* mutant strain.

### 4.3 Phenotypic analysis of *fliR*Neo mutant:

The phenotype of the *fliR* mutant was analysed under light and electron microscopy, and by swarm plates. The mutant strain (*fliR*Neo) and wild type C58C<sup>1</sup> were each inoculated into the centre of 0.2% LB swarm plates. The plates were incubated at 28 °C typically for a period of 2-3 days. The swarms formed by the mutant strain were very small, compared to those of the wild type strain, (figure 4.3.2) indicating a possible defect in the motility machinery.

Further phenotypic analysis, using microscopic observation, was used to find out whether the difference seen in the swarming size of the mutant was due to a change in the construction or action of the flagellar machinery of the cells, compared to wild type *A.tumefaciens*. Visualisation by light microscopy showed that the *fliR* mutant was non-motile, confirming the result obtained from swarm analysis. Further analysis of the mutant cells under the electron microscope showed that the cells were non-flagellated (see figure 4.3.3), therefore explaining the non-motile phenotype seen in previous tests.

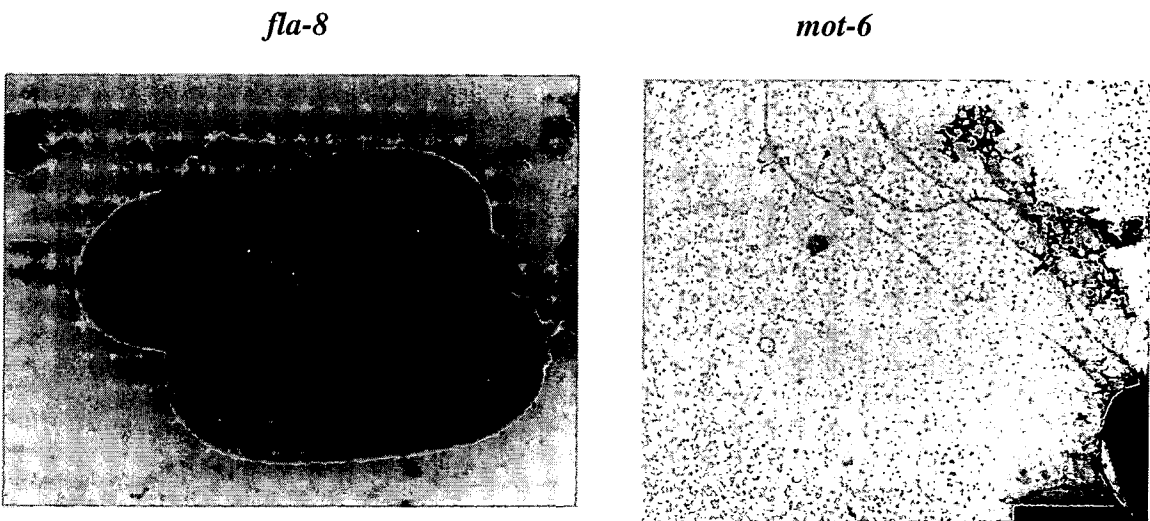
### 4.4 Discussion:

Previous work producing mutations by insertion of the transposon Tn5 into chromosomal DNA of *A. tumefaciens* then created two non-motile mutants, *fla-8* and *mot-6* (180). The two mutants share an identical Tn5 insertion site. The *fliR* mutant produced in this work showed the same phenotype as the non-motile mutants *fla-8* and *mot-6*. This was not surprising as the two methods of mutation, disrupted a region of chromosomal DNA which spans *fliR*, a member of the flagellum-specific export pathway. A similar phenotype has been seen in other bacteria, such as *C. crescentus* (238) and *R. sphaeroides* (75) when *fliR* was mutated. It has been shown that defects in any of the flagellum-specific export genes causes delivery of flagellar proteins to be blocked and assembly of flagella to be affected as a result. For instance, mutations in *flhA* affect the translocation of a number of the *E. coli* flagellar proteins (219). Mutation in *flhA* created by insertion of aphA-3 cassette, produced a non-motile and non-flagellated phenotype in *Y. enterocolitica* (4). A similar phenotype was seen in *flhA* mutation in *B. subtilis* (36). *fliQ* mutant in *C. crescentus* produced a non-motile, non-flagellated phenotype (238). A site-specific mutant of *fliI*, which was performed in this laboratory, resulted in non-motile phenotype with no visible flagella in *A. tumefaciens*. Interestingly, it has been observed that mutation on some components of



flagellum export apparatus, not only disrupt flagellar biogenesis but also normal cell division as is seen in mutants of *fliR* and *fliQ* in *C. crescentus* (238), and *flhA* in *Proteus mirabilis* (80). These observation leads to suggestion that some classII flagellum-specific export proteins apparatus genes may share, at least partially, their regulatory pathway with other cell cycle events.

The phenotype seen in the *fliR*Neo mutant was similar to that of *fla-8* but not *mot-6*, in terms of having flagella (see figure 4.4.1). Considering the fact that the phenotype of the most *fliR* mutant is non-flagellated, and the two mutants share the same Tn5 insertion site, raises a question about the source of variation between the two mutants. One explanation for this at this stage of experiment, is that because the transposon Tn5 insertion site was toward one end of the *fliR* gene in *A. tumefaciens*, sufficient transcription and translation may occur to produce a partially functional FliR protein. This could allow exportation of flagellar proteins to proceed and fragile



**Figure 4.4.1 showing *fla-8*, which is unflagellated and *mot-6* mutant, which is flagellated**

flagella to be produced under certain conditions. In this case both mutants should have flagella, however it is possible that the fragile flagella present in the *fla-8* mutant were lost in the media while being prepared for observation under the electron

microscope, and the flagella in the *mot-6* mutant may remain untouched due to different conditions. In other words, the two mutants could in fact be only one mutant. This hypothesis could be true if Tn5 transposon was inserted toward the end of the gene, however Tn5 was found to locate at the beginning of the *fliR*. in this case the gene is not transcribed to produce FliR protein necessary for flagellar synthesis. It is possible that during the experiment and mutant process some uncontrolled error was made and *mot-6*, a wild-type bacteria with a normal flagella showed mutant phenotype and assigned as a mutant.

## **5. Conclusion**

## 5. Conclusion:

The evidence obtained so far, suggests that *A. tumefaciens* like other bacteria, possess a flagellum-specific export apparatus which is used to deliver the extracellular protein units of the flagella that do not contain a cleavable leader signal peptide. Some components of this pathway were identified in the past, from work carried out in this laboratory (see section 1.4.3). Two more potential components of the flagellum-specific export system were found in this work. At present, it appears that *A. tumefaciens* contains six of the ten common components of the system, defined by Minamino *et al* (138). However, three other common components, *fliO*, *fliQ*, *fliH* and other members of the pathway, such as chaperones, which may be specific to *A. tumefaciens* or the  $\alpha$ -subgroup of proteobacteria, have not yet been identified.

The protein products of two genes (*flhA* and *fliR*) show similarity to both the flagellum-specific export pathway and typeIII proteins. The homology between *Agrobacterium* FlhA and FliR, and their counterparts in the flagellum-specific export system in other bacteria was higher than similarity to typeIII proteins. This level of similarity may count as evidence indicating that the two genes are flagellar-related genes, rather than typeIII, which it would be unlikely to find in *Agrobacterium*. The given data cannot state whether the two genes belong to an operon upstream of *flhA* or whether they form an operon with genes found downstream of *fliR*. Further analysis of the region immediately upstream of *flhA* is required to determine whether there are more flagellar-related genes, a promoter or a terminator in this region. If any promoter-like sequence was found, transcriptional fusion experiments using *lacZ*, could be carried out to analyse the promoter function. In light of what is known, it would be possible to determine the boundary of the putative operon, as well as any other putative transcription signals within it, and find the relationship between them and the rest of the flagellum-specific export genes. Northern blot analysis could also be used to identify the organisation of the genes in the operon.

In order to study the function of *flhA* in *A. tumefaciens* in more detail, mutation analysis should be made on the *flhA*. This would reveal the effect of disrupting *flhA* in flagella biogenesis and then motility. It has been shown that inactivation of the *flhA* gene leads to non-motile phenotype in other bacteria. FlhA in addition to its protein export activity has a regulatory function too. *flhA* in *C. crescentus* is required for

expression of several flagellar gene, including *flaN*, *flbG*, *flgL*, and *flgK*. In a non-motile mutant of *Proteus* lacking the *flhA* gene, expression of class II and III flagella genes is suppressed. Surprisingly, the *flhA* mutant did not elongate beyond vegetative cell length (80) which has been seen in the *fliR* mutant (see section 4.3)

Mutation in *A. tumefaciens fliR* gene result in a non-flagellated cell as has been observed in many other bacteria. An attempt to complement the *flhA* and *fliR* mutants by other bacterial counterpart genes would show whether the proteins have functional homology as well as sequence similarities. It has been known that *fliR* is activated early in the cell cycle (238). Vegetative growth of both mutants could be measured under electron microscopy to find the effect of mutation in cell division in *Agrobacterium*.

At present we have no idea as to the location and function of the gene products (FlhA and FliR) in *Agrobacterium*. It has been suggested that in *S. typhimurium* the membrane-associated components of flagellum-specific export system (FlhB, FlhA, FliO, FliQ and FliR) reside in the central pore within the MS-ring of the flagellar basal body (137). Whether this suggestion is correct for *Agrobacterium* needs to be investigated. To find the location of FlhA and FliR, immunoblotting analysis (antibodies/ western blotting) of purified inner and outer membrane fractions could be used. In this experiment the FlhA and FliR protein would be fused to a fragment of  $\beta$ -lactamase, which lacks its signal peptide for membrane translocation.

Besides the two components of the flagellum-specific export genes, a chemotaxis gene, *cheL*, along with four unknown *orfs* were also found in this work. The function of *cheL* remains to be determined, however mutation in the region containing *cheL*, resulted in a flagellated strain with a wild type motility phenotype (180). A *cheL* mutant in *C. crescentus* does not respond to chemotactic substrates, indicating a defect in the chemotaxis system (101). The same result was obtained in *A. tumefaciens* for the *che-2* mutant (180) which was mutated in a region spanning *cheL*. However, methylation is reduced in the *A. tumefaciens che-2* mutant (178), whereas the methylation in this type of mutation would be increased. It has been shown that mutations in other *che* genes in *A. tumefaciens* produces impaired chemotactic capabilities with a wild type pattern of motility (228). This leads to the suggestion that there may be an additional level of control of the flagellar motor, independent to that from the chemotaxis system. The function of protein product of *cheL* in the chemosensory system is not yet known. Mutation analysis, followed by chemotaxis assays using specific chemotractants, will

show the effect of *cheL* in the chemosensory network. The effect of this gene on motility of *A. tumefaciens* could be studied using a computer-linked tracking system that analyses statistically the motion of the bacteria, comparing it with wild type motion.

The four other *orfs* identified did not have homology to any previously known flagella or chemotaxis genes. It may be because the gene homologues to the *orfs* have not yet been identified in other bacteria, or because they are specific to *Agrobacterium*. It has been shown that bacteria belonging to  $\alpha$ -subgroup of proteobacteria are equipped with a number of unique genes (57) along with known gene homologues that are different from the enterobacteria. Mutation in the *orfs* could be used, in the first place, to study the effect of the *orfs* in motility or chemotaxis. Primer extension experiments could be carried out to show whether they are transcribed together, along with the other genes found in this fragment. Protein expression studies would be then required to show that the protein products of the *orfs* are expressed. Any effects on motility from mutagenesis or over expression may also help to determine the roles of the putative proteins. If any of the proteins were expressed, the immunoblotting could be carried out to localise their position in the cell, thus giving a possible insight into its function.

There is not enough information to explain how the flagellar export gene homologues, *cheL* and the four *orfs* are transcribed and regulated. A good transcription termination sequence was found downstream of *orfW*, it is therefore plausible to suggest that *flhA*, *fliR*, *cheL* and *orfXYW*, but not *orfZ*, form an operon or are part of an operon containing more genes upstream of *flhA*. Further experimental work will be necessary to resolve whether this speculation is correct.

The three cosmids, pDUB1900, pDUB1911 and pDUB1905 contain several flagellar and chemotaxis genes (57, 228). An experiment was carried out using hybridisation of fragments of these cosmids to cosmids from *S. meliloti* containing similar genes, (pRZ1, pRZ2 and pRZ4), which are known to overlap each other (55). The work showed that the three *Agrobacterium* cosmids to be close, however they do not overlap. It may be useful to identify a cosmid from an *A. tumefaciens* library that overlaps the cosmids pDUB1900, pDUB1911 and pDUB1905. Pulsed-field gel electrophoresis (PFGE), in combination with Southern hybridisation, could be used as a powerful tool to unify the sequences obtained from the three cosmids in *A. tumefaciens*. This would explain the relationship between the genes, whether they are clustered into one large fragment, or whether there are intergenic gaps between the clusters. It would

also give the location of the cosmids on the chromosome and would determine the exact distance between their *Bam*HI sites.

The results obtained in this work provide a lot of scope for further work, particularly on the flagellum-specific export system, which is an important system without which, no flagella can be built. It is thought that this transport protein machinery has general components that are common in all bacteria, and members that are specific for each organism. It is known that *A. tumefaciens* contains most of the general components, however the specific members should be identified. One possibility is that the ORFs found in this work could be the specific members of the typeIII flagellum-specific export system. The location, regulation and function of various components of the system also need to be investigated. The flagellar proteins that each of the components deliver could be a subject of research. The way that the system works and the interaction between the components to transport proteins also need to be found. Although, in recent years, much is known about function of this system in *S. typhimurium*, however the flagellum export system in *A. tumefaciens* could work in different way, since the structure and organisation of flagellar proteins of the two organisms are different.

Investigation of the flagellum-specific transport components could be extended to find out if the homologue components are functionally conserved among other bacteria.

Homologies between flagellum export system and typeIII secretion system suggests that two systems could be related functionally too. *A. tumefaciens* is about only Gram-negative bacteria that does not possess the typeIII secretion system. Thus, it is interesting to investigate the interaction between flagellum export system and typeIII secretion system in the sense that proteins (effectors in typeIII, and flagellar proteins in flagellum export systems) are translocated by one system could be exported or even secreted intracellularly by another system. It has been shown that the corresponding components of the typeIII system in different bacteria are not interchangeable. However, the existence of interaction between different typeIII export/secretion systems has been demonstrated in many experiments. In one occasion (52) the *pcrV* gene from *P. aeruginosa* show to complement an *lcrV* mutation in *Y. pseudotuberculosis*, and ExoS protein from *P. aeruginosa* and IpaB from *S. flexneri* could be secreted by *Y. pseudotuberculosis* typeIII system. *S. typhimurium* can also translocate YopE (52). It has been found that the *Y. enterocolitica* unable to synthesise flagellum could not also

produce several secreted proteins. In another experiment, *mopE*, a typeIII secretion system component, mutant did not produce flagella (142). A more exciting experiment was that which showed the YpIA, a virulence-associated phospholipase, which is required for *Yersinia* pathogenesis was exported by the flagellar export apparatus (232). This suggests that flagellum export machinery may have a role in the secretion of proteins involved in bacterial-host interaction. Therefore, it would be useful to determine the functional conservation between corresponding proteins of the flagellum-specific export system and the typeIII pathway. This would enable a more accurate look into the export machinery containing a superfamily of proteins that are used in a variety of ways to control and enable the process of export or assembly, or both, of proteins that lie beyond the cell membrane.

The interaction between the flagellum-specific export pathway and pathogenicity in *Agrobacterium* should also be considered. The motile mutant could be inoculated directly into test plants in order to assess their tumourgenicity. However, this method might not show the requirement of motility, as prior to attachment for the bacteria to move toward their target. It has been shown that mutation in any of the flagellum export components leads to a non-motile phenotype (75, 127, 136, 137). The role of motility in virulence has been investigated in different bacteria. It has been reported that motility mutant, non-flagellated bacteria, are affected in their virulence because of aberrations in their host adherence, invasion mechanism or other unknown factors (42). The effect of motility in *Agrobacterium* pathogenicity is not clear. Results obtained from studies on *fla*- mutants in *A. tumefaciens*, including mutations in some of the flagellum-specific export genes, showed that there were no differences between the parent and mutant in the ability to bind suspension cells and cause tumours (180). However, Chesnokova, O. *et al.* (42), have reported that virulence in non-motile, bald strains, which were generated by deletion of *flaABC*, was significantly (in statistical terms) below that of the parental wild type. Some results suggest that in bacteria that contain both systems of protein export, (The flagellum-specific export pathway and the typeIII export system) pathogenicity could be affected by a motility disruption. There is evidence showing that mutation in a typeIII protein, MopE (a FliR homologue), in *E. carotovora*, affected assembly of the flagella and lead to a non-motile phenotype (142). These results may suggest that the virulence factor export pathway may also be used for export of flagellar components, and vice versa. It is possible that the two systems share a common point in their routes. However, in *Agrobacterium* a different mechanism of



pathogenicity, which involves transfer of T-DNA to host cells, is used, and a different pathway, typeIV, is responsible for transfer of the T-DNA. Therefore disruption of the flagellum-export proteins should not affect T-DNA transfer and then tumourigenicity. However much more work is required to be done to fully understand the biology, physiology and biochemistry of the flagellum-specific export system in *A. tumefaciens*.

It should now possible to study biogenesis of flagella, motility, and in particular, the role of flagellum-specific export system in more detail in *A. tumefaciens*. This should ultimately lead to a greater understanding of *A. tumefaciens* biology in general, increasing the knowledge of *Agrobacterium* as a genetic tool and a plant pathogen.

**6. References:**

(

1. **Adler, J.** 1975. Chemotaxis in bacteria Annu. Rev. Biochem. **44**:341-356.
2. **Adler, J.** 1973. A method for measuring chemotaxis and use of the method to determine optimum conditions for chemotaxis by *Escherichia coli*. J. Gen. Microbiol. **74**:77-91.
3. **Aizawa, S.-i., C. S. Harwood, and R. J. Kadner** 2000. signaling components in bacterial locomotion and sensory reception Journal of bacteriology. **182**:1459-1471.
4. **Allen, F., A. Allaoui, A. Campos, A. Van Elsen, G. R. Cornelis, and A. Alex Bollen** 1997. Flagellar *flhA*, *flhB* and *flhE* genes, organized in an operon, cluster upstream from the *inv* locus in *Yersinia enterocolitica* Microbiology (Reading). **143**:3461-3471.
5. **Anderson, D. M., D. E. Fouts, A. Collmor, and O. Scheneewind** 1999. Reciprocal secretion of proteins by the bacterial typeIII machines of plant and animal pathogens syggests universal recognition of mRNA targeting signals Microbiology. **96**:12839-12843.
6. **Anderson, D. M., and O. Schneewind** 1997. A mRNA signal for the typeIII secretion of Yop proteins by *Yersinia enterocolitica*. Science. **278**:1140-1143.
7. **Armitage, J. P.** 1992. Behavioral responses in bacteria. Annu. Rev. Physiol. **54**:683-714.
8. **Armitage, J. P.** 1997. Behavioural Responses of Bacteria to Light and Oxygen. Arch. Microbiol. **168**:249 - 261.
9. **Armitage, J. P.** 1993. Methylation-independent behavioural responses in bacteria, p. 43-65. In J. Kurjan, and B. L. Taylor (eds), Signal Transduction: Prokaryotic and Simple Eukaryotic Systems. Academic Press, Inc., San Diego.
10. **Armitage, J. P., W. A. Havelka, and R. E. Sockett** 1990. Methylation-independent taxis in bacteria, p. 177-197. In J. P. Armitage, and J. M. Lackie (eds), Molecular Biology of the Chemotactic Response., vol. 46. Cambridge University Press, Cambridge.
11. **Armitage, J. P., and R. Schmitt** 1997. Bacterial Chemotaxis: *Rhodobacter sphaeroides* and *Sinorhizobium meliloti* - Variations on a Theme. Microbiology. **143**:3671 - 3682.
12. **Ashby, A. M.** 1988. Ph.D. University of Durham.

13. **Ashby, A. M., M. D. Watson, G. J. Loake, and C. H. Shaw** 1988. Ti Plasmid-Specified Chemotaxis of *Agrobacterium tumefaciens* C58C1 Toward *vir*-Inducing Phenolic Compounds and Soluble Factors From Monocotyledonous and Dicotyledonous Plants. **170**:4181 - 4187.
14. **Ashby, A. M., M. D. Watson, G. J. Loake, and C. H. Shaw** 1988. Ti-plasmid specified chemotaxis of *Agrobacterium tumefaciens* C58C1 towards *vir*-inducing phenolic compounds and soluble factors from monocotyledonous and dicotyledonous plants. *J. Bacteriol.* **170**:4181-4187.
15. **Barak, R., and M. Eisenbach** 1992. Correlation between phosphorylation of the chemotaxis protein CheY and its activity at the flagellar motor. *Biochem.* **31**:1821-1826.
16. **Barak, R., and M. Eisenbach** 1996. Regulation of interaction between signaling protein Che Y and flagellar motor during bacterial chemotaxis., p. 137-158, *Current Topics in Cellular Regulation.*, vol. 34. Academic Press, Inc.
17. **Beijersbergen, A., A. Den Dulk-Ras, R. A. Schilperoort, and P. J. J. Hooykaas** 1992. Conjugative Transfer by the Virulence System of *Agrobacterium tumefaciens*. *Science.* **256**:1324-1327.
18. **Bellsolell, L., P. Cronet, M. Majplero, L. Serrano, and M. Coll** 1996. The 3-dimensional structure of 2 mutants of the signal-transduction protein CheY suggest its molecular activation mechanism. *Journal of Molecular Biology.* **257**:116 - 128.
19. **Bennett, J. C. Q., and C. Colin Hughes** 2000. From flagellum assembly to virulence: the extended family of typeIII export chaperones *Trends in Microbiology.* **8**:202-204.
20. **Berg, H. C., and D. A. Brown** 1972. Chemotaxis in *Escherichia coli* analysed by three dimensional tracking. *Nature.* **239**:500-504.
21. **Berry, R. M., and J. P. Armitage** (eds.) 1999 the bacterial flagella motor, vol. 41. academic press.
22. **Bibikov, S. I., R. Brain, K. E. Rudd, and J. S. Parkinson** 1997. A signal transducer for aerotaxis in *E. coli*. *Journal of Bacteriology.* **179**:4075-4079.
23. **Bilwes, A. M., L. A. Alex, B. R. Crane, and M. I. Simon** 1999. Structure of CheA, a Signal-Transducing Histidine Kinase. *Cell.* **96**:131-141.

24. **Binns, A. N., C. E. Beaupre, and E. M. Dale** 1995. Inhibition of VirB-mediated transfer of diverse substrates from *Agrobacterium tumefaciens* by the incQ plasmid RSF1010 J. bacteriol. **177**:4890-99.
25. **Bischoff, D. S., and G. W. Ordal** 1992. *Bacillus subtilis* chemotaxis: a deviation from the *Escherichia coli* paradigm. Mol. Microbiol. **6**:23-28.
26. **Blair, D. F.** 1995. How Bacteria Sense and Swim. Annual Review of Microbiology. **49**:489-522.
27. **Blat, Y., and M. Eisenbach** 1996. Mutants with defective phosphatase activity show no phosphorylation-dependent oligomerization of CheZ. Journal of Biological Chemistry. **271**:1232-1236.
28. **Blat, Y., and M. Eisenbach** 1994. Phosphorylation-Dependant Binding of the Chemotaxis Signal Molecule CheY to its Phosphatase, CheZ. Biochemistry. **33**:902-906.
29. **Bourret, R. B., K. A. Borkovich, and M. Simon** 1991. signal transduction pathways involving protein phosphorylation in prokaryotes Ann. Rev. Biochem. **60**:401-444.
30. **Bourret, R. B., J. F. Hess, K. A. Borkovich, A. A. Pakula, and M. Simon** 1989. Protein phosphorylation in chemotaxis and two component regulatory systems of bacteria J. Biol. Chem. **264**:7085-7088.
31. **Bouzar, H., D. Ouadah, Z. Krimi, J. B. Jones, M. Trovato, A. Petit, and Y. Desaux** 1993. Correlative association between resident plasmids and the host chromosome in a diverse *Agrobacterium* soil population Appl. Environ. Microbiol. **59**:1310-1317.
32. **Bravo-Angel, A. M., B. Hohn, and B. Tinland** 1998. The Omega Sequence of VirD2 is Important but not Essential for efficient Transfer of T-DNA by *Agrobacterium tumefaciens*. Molecular Plant-Microbe Interactions. **11**:57 - 63.
33. **Breyne, P., and G. Gheysen** 1992. Effect of T-DNA configuration on transgene expression Mol. Gen. Genet. **235**:389-396.
34. **Brown, A. P.** 1992. Ph.D. University of Durham.
35. **Burns, D. L.** 1999. Biochemistry of typeIV secretion. **2**:25-29.
36. **Carpenter, P. B., and G. W. Ordal** 1993. bacillus subtilis FlhA: a flagellar protein related to a new family of signal-transducing receptors molecular microbiology. **7**:735-743.

37. **Carpenter, p. B., A. R. Zuberi, and G. W. Ordal** 1993. *Bacillus subtilis* flagellar proteins FliP, FliQ, FliR and FlhB are related to *Shigella flexneri* virulence factors Gene. **137**:243-245.
38. **Chang, C., and R. C. Stewart** 1998. The Two Component System. Plant Physiology. **117**:723-731.
39. **Chang, C., and R. C. Stewart** 1998. The Two-Component System Plant Physiology. **117**:723-731.
40. **Chang, C., and S. C. Winans** 1992. Functional roles assigned to the periplasmic, linker and receiver domains of the *Agrobacterium tumefaciens* VirA protein J. Bacteriol. **174**:7033-7039.
41. **Chervitz, S. A., and J. J. Falke** 1996. Molecular mechanism of transmembrane signaling by the aspartate receptor: A model. Proc. Natl. Acad. Sci. (USA). **93**:2545 - 2550.
42. **Chesnokova, G. A., J. B. Coutinho, I. H. Khan, M. S. Mikhail, and C. I. Kado** 1997. Characterization of Flagella Genes of *Agrobacterium tumefaciens*, and the Effect of a Bald Strain on Virulence. Molecular Microbiology. **23**:579-590.
43. **Chiang, P. K., R. K. Gordon, J. Tal, G. C. Zeng, B. P. Doctor, K. Pardhasaradhi, and P. P. McCann** 1996. S-Adenosylmethionine and Methylation. FASEB Journal. **10**:471-480.
44. **christie, P. J.** 1997. *Agrobacterium tumefaciens* T-complex transport apparatus: a paradigm for a new family of multifunctional transports in eubacteria. J. Bacteriol. **179**:3058-3094.
45. **Christie, P. J., and J. P. Vogel** 2000. Bacterial typeIV secretion: conjugation systems adapted to deliver effector molecules to host cells. **8**:354-360.
46. **Christie, P. J., and J. P. Vogel** 2000. Bacterial typeIV secretion: conjugation systems adapted to deliver effector molecules to host cells Trends in Microbiology. **8**:354-360.
47. **Citovsky, V., M. Lie Wong, and P. C. Zambryski** 1989. Co-operative interaction of *Agrobacterium* VirE2 protein with single-stranded DNA: implications for the T-DNA transfer process Proc. Natl. Acad. Sci. USA. **86**:1193-1197.

48. **Citovsky, V., D. Warnick, and P. Zambryski** 1994. Nuclear Import of *Agrobacterium* VirD2 and VirE2 proteins in Maize and Tobacco. *Proc. Natl. Acad. Sci. USA.* **91**:3210 - 3214.
49. **Citovsky, V., J. Zupan, D. Warnick, and P. Zambryski** 1992. Nuclear Localization of *Agrobacterium* VirE2 Protein in Plant Cells. *Science.* **256**:1802 - 1805.
50. **Conley, M. P., A. J. Wolfe, D. F. Blair, and H. C. Berg** 1989. Both CheA and CheW are Required for Reconstitution of Chemotactic Signalling in *Escherichia coli*. *Journal of Bacteriology.* **171**:5190-5193.
51. **Cooley, M. B., M. R. D'Souza, and C. I. Kado** 1991. The *virC* and *virD* operons of the *Agrobacterium* Ti-plasmid are regulated by the *ros* chromosomal gene: analysis of the cloned *ros* gene. *J. Bacteriol.* **173**:2608-2616.
52. **Cornelis, G. R.** 2000. TypeIII secretion: a bacterial device for close combat with cells of their eukaryotic host *Phil. Trans. R. lond. B.* **355**:681-693.
53. **Danielson, M. A., R. B. Bass, and J. J. Falke** 1997. Cysteine nad Disulfide Scanning Reveals a regulatory  $\alpha$ - Helix in the Cytoplasmic Domain of the Aspartate Receptor. *Journal of Biological Chemistry.* **272**:32878 - 32888.
54. **Davey, M. R., I. S. Curtis, K. M. A. Gartland, and J. B. Power** 1994. *Agrobacterium*-Induced Crown Gall and Hairy Root Diseases: Their biology and Application to Plant Genetic Engineering. *In* M. A. J. Williams (ed.), *Plant Galls*, vol. 49. Clarendon Press, Oxford.
55. **Deakin, W. J.** 1994. PhD Thesis. University of Durham.
56. **Deakin, W. J., C. S. M. Furniss, V. E. Parker, and C. H. Shaw** 1997. Isolation and Characterisation of a Linked Cluster of Genes From *Agrobacterium tumefaciens* Encoding Proteins Involved in Flagellar Basal-body Structure. *Gene.* **189**:135-137.
57. **Deakin, W. J., V. E. Parker, E. L. Wright, K. J. Ashcroft, G. J. Loake, and C. H. Shaw** 1999. *Agrobacterium tumefaciens* possesses a Fourth Flagellin Gene Located in a Large Gene Cluster concerned with Flagellar Structure, Assembly and Motility. *Microbiology-UK.* **145**:1397-1407.
58. **Deng, W., L. Chen, W. T. Peng, X. Liang, S. Sekiguchi, M. P. Gordon, L. Comai, and E. W. Nester** 1999. VirE1 is a specific molecular chaperone for

the exported single-stranded-DNA-binding protein VirE2 in *Agrobacterium*. **31**:1795-1807.

59. **Deng, W., L. Chen, W.-T. Peng, X. Liang, S. Sekiguchi, M. P. Gordon, L. Comai, and E. W. Nester** 1999. VirE1 is a specific molecular chaperone for the exported single-stranded-DNA-binding protein VirE2 in *Agrobacterium* Molecular Microbiology. **31**:1795-1807.
60. **Deng, W., L. Chen, D. W. Wood, T. Metcalfe, X. Liang, M. P. Gordon, L. Comai, and E. W. Nester** 1998. *Agrobacterium* VirD2 Protein Interacts with Plant Host Cyclophilins. Proc. Natl. Acad. Sci. USA. **95**:7040-7045.
61. **Dessaux, Y., J. Tempe, and S. K. Farrand** 1987. Genetic analysis of mannityl opine catabolism in octopine-type *Agrobacterium tumefaciens* strain 15955. Mol. Gen. Genet. **208**:301-308.
62. **Ditta, G., S. Stanfield, D. Corbin, and D. R. Helsinki** 1980. A broad host range DNA cloning system for Gram-negative bacteria: Construction of a gene bank of *Rhizobium meliloti*. Proc. Natl. Acad. Sci. USA. **77**:7347-7351.
63. **Djordjevic, S., and A. M. Stock** 1998. Structural Analysis of Bacterial Chemotaxis Proteins: Components of a Dynamic Signaling System Journal of Structural Biology. **124**:189-200.
64. **Dombek, P., and W. Ream** 1997. Functional Domains of *Agrobacterium tumefaciens* Single-Stranded DNA-Binding Protein VirE2. Journal of Bacteriology. **179**:1165 - 1173.
65. **Duban, M. E., K. Lee, and D. G. Lynn** 1993. Strategies in Pathogenesis: Mechanistic Specificity in the Detection of Generic Signals. Molecular Microbiology. **7**:637-645.
66. **Eisenbach, M.** 1996. Control of bacterial chemotaxis. **20**:903-910.
67. **Escudero, J., and B. Hohn** 1997. Transfer and Integration of T-DNA Without Cell Injury in the Host Plant. The Plant Cell. **9**:2135-2142.
68. **Fabrice, D., M. Duckely, P. Pelczar, P. Van Gelder, and B. Hohn** 2001. An *agrobacterium* VirE2 channel for transferred-DNA transport into plant cells Proc. Natl. Acad. Sci. (USA). **98**:485-490.
69. **Falke, J. J., R. B. Bass, S. L. Butler, S. A. Chervitz, and M. A. Danielson** 1997. The Two-Component Signalling Pathway of Bacterial Chemotaxis: A Molecular View of Signal Transduction by Receptors, Kinases, and Adaptation Enzymes., Annual Review of Cell and Developmental Biology.



70. **Farrand, S. K., I. Hwang, and D. M. Cook** 1996. The tra Region of the Nopaline-Type Ti Plasmid Is a Chimera with Elements Related to the Transfer Systems of RSF1010, RP4 and F. *Journal of Bacteriology*. **178**:4233 - 4247.
71. **Fernandez, L. A., and J. Berenguer** 2000. secretion and assembly of regular surface structures in gram-negative bacteria *FEMS Microbiology Reviews*. **24**:21-44.
72. **Filichkin, S. A., and S. B. Gelvin** 1993. Formation of a Putative Relaxation Intermediate During T-DNA Processing Directed by the *Agrobacterium tumefaciens* VirD1, D2 endonuclease. *Molecular Microbiology*. **8**:915 - 926.
73. **Frieberg, C., R. Felly, A. Bairoch, W. J. Broughton, A. Rosenthal, and X. Perret** 1997. Molecular basis of symbiosis between *Rhizobium* and legumes. *Nature*. **387**:394-401.
74. **Galán, J. E., C. Ginocchio, and P. Costeas** 1992. Molecular and functional characterization of the *Salmonella* invasion gene *invA*: homology of InvA to members of a new protein family. *J. Bacteriol.* **174**:4338-4349.
75. **Garcia, N., A. Campos, A. Osorio, S. Poggio, B. Gonzalez-Pedrajo, L. Camarena, and G. Dreyfus** 1998. The Flagellar Switch Genes *fliM* and *fliN* of *Rhodobacter sphaeroides* are Contained in a Large Flagellar Gene Cluster. *Journal of Bacteriology*. **180**:3978-3982.
76. **Gegner, J. A., and F. W. Dahlquist** 1991. Signal transduction in bacteria: CheW forms a reversible complex with the protein kinase CheA. *Proc. Natl. Acad. Sci. USA*. **88**:750-754.
77. **Gelvin, S. B.** 1998. *Agrobacterium* VirE2 Proteins Can Form a Complex with T Strands in the Plant Cytoplasm. *Journal of Bacteriology*. **180**:4300-4302.
78. **Grebe, T. W., and J. Stock** 1998. Bacterial Chemotaxis: The Five Sensors of a Bacterium. *Current Biology*. **8**:R154 - R157.
79. **Greck, M., J. Platzer, V. Sourjik, and R. Schmitt** 1995. Analysis of a chemotaxis operon in *Rhizobium meliloti*. *Molecular Microbiology*. **15**:989 - 1000.
80. **Gugi, D., M. J. Bailey, C. Allison, and C. Hughes** 1995. Requirement for FlhA in assembly and swarm-cell differentiation by *proteus mirabilis*. *Molecular microbiology*. **15**:761-769.

81. **Hamblin, P. A., N. A. Bourne, and J. P. Armitage** 1997. Characterization of the Chemotaxis Protein CheW from *Rhodobacter sphaeroides* and its Effect on the Behaviour of *Escherichia coli*. *Molecular Microbiology*. **24**:41 - 51.
82. **Hamblin, P. A., B. A. Maguire, R. N. Grishanin, and J. P. Armitage** 1997. Evidence for two Chemosensory Pathways in *Rhodobacter sphaeroides*. *Molecular Microbiology*. **26**:1083 - 1096.
83. **Han, D. C., and S. C. Winans** 1994. A Mutation in the Receiver Domain of the *Agrobacterium tumefaciens* Transcriptional Regulator VirG Increases its Affinity for Operator DNA. *Molecular Microbiology*. **12**:23-30.
84. **Herrera-Estrella, A., M. Van Montagu, and K. Wang** 1990. A bacterial peptide acting as a plant nuclear targeting signal: the amino-terminal portion of *Agrobacterium* VirD2 protein directs  $\beta$ -galactosidase fusion protein into tobacco nuclei. *Proc. Natl. Acad. Sci. USA*. **87**:9534-9537.
85. **Hess, J. F., R. B. Bourret, and M. I. Simon** 1988. Histidine phosphorylation and phosphoryl group transfer in bacterial chemotaxis. *Nature*. **336**:139-143.
86. **Holsters, M., R. Villaroel, J. Gielen, J. Seurinck, H. de Greve, M. Van Montagu, and J. Schell** 1983. An analysis of the boundaries of the octopine TL-DNA in tumours induced by *Agrobacterium tumefaciens* *Mol. Gen. Genet.* **190**:35-41.
87. **Homma, M., T. Iino, K. Kutsukake, and S. Yamaguchi** 1986. *In vitro* reconstitution of flagellar filaments onto hooks of filamentless mutants of *Salmonella typhimurium* by addition of hook-associated proteins. *Proc. Natl. Acad. Sci. USA*. **83**:6169-6173.
88. **Hooykaas, P. J. J., and A. G. M. Beijersbergen** 1994. The virulence system of *Agrobacterium tumefaciens*. **32**:157-179.
89. **Hooykaas, P. J. J., and A. G. M. Beijersbergen** 1994. The virulence system of *Agrobacterium tumefaciens*. *Annual Review of Phytopathology*. **32**:157-179.
90. **Hooykaas, P. J. J., P. M. Klapwijk, M. P. Nuti, R. A. Schilperoort, and A. Rörsch** 1977. Transfer of the *Agrobacterium tumefaciens* Ti-plasmid to avirulent *Agrobacteria* and to *Rhizobium ex planta* *J. Gen. Microbiol.* **98**:477-484.

91. **Howie, W. J., R. J. Cook, and D. M. Weller** 1987. Effects of soil matrix potential and cell motility on wheat root colonization by fluorescent pseudomonads suppressive to take-all Phytopathology. **77**:286-292.
92. **Hueck, C. J.** 1998. TypeIII protein secretion systems in bacterial pathogens of animals and plants Microbiology and Molecular Biology Reviews. **62**:379-433.
93. **Jin, S., R. K. Prusti, T. Poitsch, R. G. Ankenbauer, and E. W. Nester** 1990. Phosphorylation of the VirG protein of *Agrobacterium tumefaciens* by the autophosphorylated VirA protein: essential role in biological activity of VirG J. Bacteriol. **172**:4945-4950.
94. **Jin, S., T. Roitsch, R. G. Ankenbauer, M. P. Gordon, and E. W. Nester** 1990. The VirA protein of *Agrobacterium tumefaciens* is autophosphorylated and is essential for *vir* gene induction J. Bacteriol. **172**:525-530.
95. **Johnson, M. S., E. H. Rowsell, and B. L. Taylor** 1995. Investigation of transphosphorylation between chemotaxis proteins and the phosphoenolpyruvate : sugar phosphotransferase system. FEBS Letters. **374**:161 - 164.
96. **Jumas-Bilak, e., S. Michaux-Charachon, G. Bourg, M. Ramuz, and A. Allardet-Servent** 1998. Unconventional Genomic Organisation in the Alpha Subgroup of the Proteobacteria. **180**:2749-2755.
97. **Kerr, A.** 1992. The genus *Agrobacterium*, p. 2214-2235. In Balows, Trüper, Dworkin, Harder, and Schleifer (eds), The Prokaryotes. Springer-Verlag, New york.
98. **Kerr, A.** 1974. Soil microbiological studies on *Agrobacterium radiobacter* and biological control of crown gall Soil Sci. **118**:168-172.
99. **Kerr, A., and C. G. Panagopoulos** 1982. Biotypes of *Agrobacterium radiobacter* var. *tumefaciens* and their biological control Phytopathol. **90**:172-179.
100. **Kerstens, K., and J. DeLey** 1984. Section 4, Family III, Genus III: *Agrobacterium* Conn 1942, p. 244-254. In N. R. Krieg (ed.), Bergey's Manual of Systematic Bacteriology. Williams and Wilkins, Baltimore.
101. **Khambaty, F. M., and B. Ely** 1992. Molecular genetics of the *flgI* region and its role in flagellum biosynthesis in *Caulobacter crescentus*. J. Bacteriol. **174**:4101-4109.

102. **Kim, H., and S. K. Farrand** 1997. Characterization of the *acc* Operon from the Nopaline-Type Ti Plasmid pTiC58, Which Encodes Utilization of Agrocinopines A and B and Susceptibility to Agrocin 84. *Journal of Bacteriology*. **179**:7559-7572.
103. **Kim, H., and S. K. Farrand** 1998. Opine Catabolic Loci from *Agrobacterium* Plasmids Confer Chemotaxis to Their Cognate Substrates. *Molecular Plant-Microbe Interactions*. **11**:131-143.
104. **Kim, K., and S. K. Farrand** 1996. Ti Plasmid-Encoded Genes Responsible for Catabolism of the Crown Gall Opine Mannopine by *Agrobacterium tumefaciens* Are Homologs of the T-Region Genes Responsible for Synthesis of this Opine by the Plant Tumour. *Journal of Bacteriology*. **178**:3275 - 3284.
105. **Kirankumar S. Mysore, B. B., Xiao-bing Deng, Nune S. Darbinian, Andrei Motchoulski, Walt Ream, and Stanton B. Gelvin** 1998. Role of the *Agrobacterium tumefaciens* VirD2 Protein in T-DNA Transfer and Integration MPMI. **11**:668-683.
106. **Kubori, T., Y. Matsushima, D. Dakamura, J. Uralil, M. Lara-Tejero, A. Sukhan, J. E. Galan, and S.-I. Aizawa** 1998. Supramolecular stacture of the *salmonella typhimurium* typeIII protein secretion system. **280**:602-605.
107. **Kubori, T., N. Shimamoto, S. Yamaguchi, K. Namba, and S.-I. Aizawa** 1992. Morphological pathway of flagellar assembly in *Salmonella typhimurium*. *J. Mol. Biol.* **226**:433-446.
108. **Lai, E.-M., and C. I. Kado** 2000. The T-pilus of *Agrobacterium tumefaciens* *Trends in Microbiology*. **8**:361-369.
109. **Lee, K., M. W. Dudley, K. M. Hess, D. G. Lynn, R. D. Joerger, and A. N. Binns** 1992. Mechanism of activation of *Agrobacterium* virulence genes: identification of phenol-binding proteins *Proc. Natl. Acad. Sci. USA*. **89**:8666-8670.
110. **Lee, L., S. B. Gelvin, and C. I. Kado** 1999. pSa Causes Oncogenic suppression of *Agrobacterium* by Inhibiting VirE2 Protein Export. *Journal of Bacteriology*. **181**:186-196.
111. **Lee, Y., S. Jin, W. Sim, and E. W. Nester** 1996. The Sensing of Plant Signal Molecules by *Agrobacterium*: Genetic Evidence for Direct Recognition of Phenolic Inducers by the VirA Protein. *Gene*. **179**.

112. **Levit, M. N., Y. Liu, and J. B. Stock** 1998. Stimulus Response Coupling in Bacterial Chemotaxis: Receptor Dimers in Signalling Arrays. *Molecular Microbiology*. **30**:459-466.
113. **Li, J., R. V. Swanson, M. I. Simon, and R. M. Weis** 1995. The response regulators CheB and CheY exhibit competitive binding to the kinase Che A. *Biochemistry*. **34**:14626 - 14636.
114. **Lin, T., and C. I. Kado** 1993. The *virD4* Gene is Required for Virulence while *virD3* and *orf5* are not Required for Virulence of *Agrobacterium tumefaciens*. *Molecular Microbiology*. **9**:803 - 812.
115. **Lippincott, B. B., and J. A. Lippincott** 1969. Bacterial attachment to a specific wound site is an essential stage in tumour initiation by *Agrobacterium tumefaciens* *J. Bacteriol.* **97**:620-628.
116. **Liu, J., and J. S. Parkinson** 1991. Genetic evidence for interaction between the CheW and Tsr proteins during chemoreceptor signaling by *Escherichia coli*. *J. Bacteriol.* **173**:4941-4951.
117. **Loake, G. J.** 1989. PhD Thesis. University of Durham.
118. **Loake, G. J., A. M. Ashby, and C. H. Shaw** 1988. Attraction of *Agrobacterium tumefaciens* C58C1 towards sugars involves a highly sensitive chemotaxis system. *J. Gen. Microbiol.* **134**:1427-1432.
119. **Lory, S.** 1998. Secretion of proteins and assembly of bacterial surface organelles: shared pathways of extracellular protein targeting *Current Opinion in Microbiology*. **1**:27-35.
120. **Lux, R., K. Jahreis, k. Bettenbrock, J. S. Parkinson, and J. W. Lengeler** 1995. Coupling the phosphotransferase system and the methyl-accepting chemotaxis signaling pathways of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*. **92**:11583 - 11587.
121. **Macnab, R. M.** 1999. The bacterial flagellum: Reversible rotary propellor and typeIII export apparatus *Journal of bacteriology*. **181**:7149-7153.
122. **Macnab, R. M.** 1996. Flagella and motility, 2nd ed, vol. 1. ASM Press, Washington, DC: American Society for Microbiology.
123. **Macnab, R. M.** 1992. Genetics and biogenesis of bacterial flagella. *Annu. Rev. Genet.* **26**:131-158.

124. **Macnab, R. M.** 1990. Genetics, structure, and assembly of the bacterial flagellum, p. 77-106. *In* J. P. Armitage, and J. M. Lackie (eds), *Biology of the Chemotactic Response*, vol. 46. Cambridge University Press, Cambridge.
125. **Macnab, R. M.** 1987. Motility and Chemotaxis, p. 732-759. *In* F. C. Neidhart, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (eds), *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology., vol. 1. American Society for Microbiology, Washington, D. C.
126. **Maddock, J. R., and L. Shapiro** 1993. Polar location of the chemoreceptor complex in the *Escherichia coli* cell. *Science*. **259**:1717-1723.
127. **Malakooti, J., B. Ely, and P. Matsumura** 1994. Molecular characterization, nucleotide sequence, and expression of the *fliO*, *fliP*, *fliQ*, and *fliR* genes of *Escherichia coli*. *J. Bacteriol.* **176**:189-197.
128. **Mantis, N. J., and S. C. Winans** 1992. The *Agrobacterium tumefaciens* *vir* gene transcriptional activator *virG* is transcriptionally induced by acid pH and other stress stimuli *J. Bacteriol.* **174**:1189-1196.
129. **Matthyse, A. G.** 1983. Role of bacterial cellulose fibrils in *Agrobacterium tumefaciens* infection. *J. Bacteriol.* **154**:906-915.
130. **Matthyse, A. G., S. White, and R. Lightfoot** 1995. Genes Required for Cellulose Synthesis in *Agrobacterium tumefaciens*. *Journal of Bacteriology*. **177**:1069-1975.
131. **McEvoy, M. M., and F. W. Dahlquist** 1997. Phosphohistidines in Bacterial Signaling. *Current Opinion in Structural Biology*. **7**:793 - 797.
132. **McEvoy, M. M., L. E. K. Muhandiram, and F. W. Dahlquist** 1996. Structure and Dynamics of a CheY-binding domain of the chemotaxis kinase CheA determined by nuclear magnetic resonance. *Biochemistry*. **35**:5633-5640.
133. **McLean, B. G., E. A. Greene, and P. C. Zambryski** 1994. Mutants of *Agrobacterium* VirA that Activate *vir* Gene Expression in the Absence of the Inducer Acetosyringone. *The Journal of Biological Chemistry*. **269**:2645-2651.
134. **McNally, D. F., and P. Matsumura** 1991. Bacterial Chemotaxis Signalling Complexes: Formation of a CheA/CheW Complex Enhances

Autophosphorylation and Affinity for CheY. Proc. Natl. Acad. Sci. USA. **88**:6269-6273.

135. **Milligan, D. L., and D. E. Koshland Jr.** 1993. Purification and characterization of the periplasmic domain of the aspartate chemoreceptor. J. Biol. Chem. **268**:19991-19997.
136. **Minamino, T., T. Iino, and k. kutsukake** 1994. molecular characterization of salmonella typhimurium flhB operon and its protein products journal of bacteriology. **176**:7630-7637.
137. **Minamino, T., and R. M. Macnab** 1999. Components of the salmonella flagellar export apparatus and classification of export substrates Journal of bacteriology. **181**:1388-1394.
138. **Minamino, T., and R. M. Macnab** 2000. INTERACTIONS AMONG COMPONENTS OF THE SALMONELLA FLAGELLR EXPORT APPARATUS AND ITS SUBSTRATES Molecular Microbiology. **35**:1052-1064.
139. **Montrone, M., M. Eisenbach, D. Oesterhelt, and W. Marwan** 1998. Regulation of Switching Frequency and Bias of the Bacterial Flagellar Motor by CheY and Fumarate., Journal of Bacteriology. **180**:3375-3380.
140. **Moreno, E.** 1998. Genome evolution within the alpha Proteobacteria: Why do some bacteria not possess plasmids and others exhibit more than one different chromosome? FEMS Microbiology Reviews. **22**:255-275.
141. **Mowbray, S. I., and M. O. J. Sandgren** 1998. chemotaxis receptors: a progress report on structure and function Journal of structural biology. **124**:257-275.
142. **Mulholland, V., J. C. D. Hinton, J. Sidebotham, I. K. Toth, L. J. Hyman, M. C. M. Pérombelon, P. J. Reeves, and G. P. C. Salmond** 1993. A pleiotropic reduced virulence (Rvi-) mutant of *Erwinia carotovora* subspecies *atroseptica* is defective in flagella assembly proteins that are conserved in plant and animal bacterial pathogens. Mol. Microbiol. **9**:343-356.
143. **Muller, J., S. Schiel, G. W. Ordal, and H. H. Saxild** 1997. functional and genetic characterization of mcpC, which encodes a third methyl-accepting chemotaxis protein in *Bacillus subtilis* Microbiology (Reading). **143**:3231-3240.

144. **Namba, K., I. Yamashita, and F. Vonderviszt** 1989. Structure of the core and central channel of bacterial flagella. *Nature*. **342**:648-654.
145. **Nambu, T., T. Minamino, R. M. Macnab, and K. Kutsukake** 1999. Peptidoglycan-hydrolyzing activity of the FlgJ protein, essential for flagellar rod formation in *Salmonella typhimurium* *Journal of bacteriology*. **181**:1555-1561.
146. **Nester, E. W., M. P. Gordon, R. M. Amasino, and M. F. Yanofsky** 1984. Crown gall - a molecular and physiological analysis *Annu. Rev. Plant Physiol.* **35**:387-413.
147. **Nester, E. W., and K. T.** 1981. plasmids specifying plant hyperplasias *Annual review of Microbiology*.
148. **New, P. B., and A. Kerr** 1972. Biological control of crown gall: field measurements and glasshouse experiments *J. Appl. Bacteriol.* **35**:279-285.
149. **Oger, P., K. Kim, R. L. Sackett, K. Piper, and S. K. Farrand** 1998. Octopine-Type Ti Plasmids Code For a Mannopine-Inducible Dominant-Negative allele of *traR*, the Quorum-Sensing Activator that Regulates Ti Plasmid Conjugal Transfer. *Molecular Microbiology*. **27**:277-288.
150. **Ohnishi, K., K. Kutsukake, H. Suzuki, and T. Iino** 1992. A novel transcriptional regulation mechanism in the flagellar regulon of *Salmonella typhimurium*: an anti-sigma factor inhibits the activity of the flagellum-specific sigma factor,  $\sigma^F$ . *Mol. Microbiol.* **6**:3149-3157.
151. **Okamoto, S., A. Toyoda-Yamamoto, K. Ito, I. Takebe, and Y. Machida** 1991. Localisation and orientation of the VirD4 protein of *Agrobacterium tumefaciens* in the cell membrane. *Mol. Gen. Genet.* **228**:24-32.
152. **Ordal, G. W., L. Márquez-Magaña, and M. J. Chamberlin** 1993. Motility and chemotaxis, p. 765-784. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (eds), *Bacillus subtilis* and other Gram-Positive Bacteria: Biochemistry, Physiology, and Molecular Genetics. American Society for Microbiology, Washington, D. C.
153. **Packer, H. L., H. Lawther, and J. P. Armitage** 1997. The *Rhodobacter sphaeroides* Flagellar Motor is a Variable-Speed Rotor. *FEBS Letters*. **409**:37-40.



154. **Pansegrau, W., F. Schoumacher, B. Hohn, and E. Lanka** 1993. Site-specific Cleavage and Joining of Single-stranded DNA by VirD2 Protein of *Agrobacterium tumefaciens* Ti Plasmids: Analogy to Bacterial Conjugation. *Proc. Natl. Acad. Sci. USA.* **90**:11538 - 11542.
155. **Parkinson, J. S.** 1993. Signal transduction schemes of bacteria. *Cell.* **73**:857-871.
156. **Parkinson, J. S., and D. F. Blair** 1993. Does *E. coli* have a nose? *Science.* **259**:1701-1702.
157. **Parkinson, J. S., and E. C. Kofoed** 1992. Communication modules in bacterial signaling proteins. *Annu. Rev. Genet.* **26**:71-112.
158. **Pazour, G. J., and A. Das** 1990. Characterisation of the VirG binding site of *Agrobacterium tumefaciens*. *Nucl. Acids Res.* **18**:6909-6913.
159. **Plano, G. V., S. S. Barve, and S. C. Straley** 1991. LcrD, a membrane-bound regulator of the *Yersinia pestis* low-calcium response. *J. Bacteriol.* **173**:7293-7303.
160. **Postma, P. W., J. W. Lengeler, and G. R. Jacobson** 1993. Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria. *Micobiol. Rev.* **57**:543-594.
161. **Pugsley, A. P.** 1993. The complete general secretory pathway in Gram-negative bacteria. *Microbiol. Rev.* **57**:50-108.
162. **Quandt, J., and M. F. Hynes** 1993. Versatile suicide vectors which allow direct selection for gene replacement in Gram-negative bacteria. *Gene.* **127**:15-21.
163. **Ramakrishnan, R., M. Schuster, and R. B. Bourret** 1998. Acetylation at Lys-92 Enhances Signaling by the Chemotaxis Response Regulator Protein CheY. *Proc. Natl. Acad. Sci. USA.* **95**:4918-4923.
164. **Rice, M. S., and F. W. Dahlquist** 1991. Sites of deamidation and methylation in Tsr, a bacterial chemotaxis sensory transducer. *J. Biol. Chem.* **266**:9746-9753.
165. **Roitsch, T., S. Jin, and E. W. Nester** 1994. The Binding Site of the Transcriptional Activator VirG from *Agrobacterium* Comprises both Conserved and Specific Nonconserved Sequences. *FEBS Letters.* **338**:127-132.

166. **Romagnoli, S., and J. P. Armitage** 1999. Roles of Chemosensory Pathways in Transient Changes in Swimming Speed of *Rhodobacter sphaeroides* Induced by Changes in Photosynthetic Electron Transport. *Journal of Bacteriology*. **181**:34-39.
167. **Romantschuk, M.** 1992. Attachment of plant pathogenic bacteria to plant surface *Annu. Rev. Phytopathol.* **30**:225-243.
168. **Rowsell, E. H., J. M. Smith, A. Wolfe, and B. L. Taylor** 1995. CheA, CheW, and CheY are Required for Chemotaxis to Oxygen and Sugars of the Phosphotransferase System in *Escherichia coli*. *Journal of Bacteriology*. **177**:6011-6014.
169. **Sambrook, J., E. F. Fritsch, and T. Maniatis** 1989. *Molecular Cloning: A laboratory manual*, 2nd ed. Cold Spring Harbour Laboratory Harbour Press, New York.
170. **sanders, D. A., B. L. Gillece-Castro, A. M. Stock, A. L. Burlingame, and D. E. Koshland** 1989. Identification of the site of phosphorylation of the chemotaxis response regulator, CheY. *J. Biol. Chem.* **264**:21770-21778.
171. **Sanders, L. A., S. Van Way, and D. A. Mullin** 1992. characterization of caulobacter crescentus flbF promoter and identification of the inferred FlbF product as a homology of the LcrD protein from a yersinia enterocolitica virulence plasmid *journal of bacteriology*. **174**.
172. **Sanders, I. A., S. V. Way, and D. A. Mullin** 1992. Characterization of caulobacter crescentus flbF promoter and identification of the inferred FlbF product as a homology of the LcrD protein from a yersinia enterocolitica virulence plasmid *journal of bacteriology*. **174**.
173. **Sanna, M. G., and M. I. Simon** 1996. Isolation and *in-vitro* characterization of CheZ suppressors for the *Escherichia coli* chemotactic response regulator mutant CheYN23D. *Journal of Biological Chemistry*. **271**:7357-7361.
174. **Sasakawa, C., T. T. Komatsu, T. Suzuki, and M. Yoshikawa** 1993. Eight genes in region 5 that form an operon are essential for invasion of epithelial cells by *shigella flexneri*. *J. bacteriol.* **175**:2334-2346.
175. **Savka, M. A., R. C. Black, A. N. Binns, and S. K. Farrand** 1996. Translocation and Exudation of Tumor Metabolites in Crown Galled Plants. *Molecular Plant-Microbe Interactions*. **9**:310-313.

176. **Scharf, B. E., K. A. Fahrner, L. Turner, and H. C. Berg** 1998. Control of Direction of Flagellar Rotation in Bacterial Chemotaxis. *Proc. Natl. Acad. Sci. USA*. **95**:201-206.
177. **Shaw, C. H.** 1996. The Molecular Basis of Rhizosphere Competence in *Agrobacterium tumefaciens*, p. 242-252. In R. W. Pickup, J. R. Saunders, and G. A. Codd (eds), *Molecular Approaches to Environmental Microbiology*. Ellis Harwood.
178. **Shaw, C. H.** 1990. Swimming against the tide: chemotaxis in *Agrobacterium*. *Bioessays*. **13**:25-29.
179. **Shaw, C. H., G. H. Carter, and M. D. Watson** 1984. A Functional Map of the Nopaline Synthase Promoter. *Nucleic Acids Research*. **12**:7831-7846.
180. **Shaw, C. H., G. J. Loake, A. P. Brown, C. S. Garrett, W. Deakin, G. Alton, M. Hall, S. A. Jones, M. O'Leary, and L. Primavesi** 1991. Isolation and characterization of behavioural mutants and genes of *Agrobacterium tumefaciens*. *J. Gen. Microbiol.* **137**:1939-1953.
181. **Sheng, J., and V. Citovsky** 1996. *Agrobacterium*-Plant Cell DNA Transport: Have Virulence Proteins, Will Travel. *The Plant Cell*. **8**:1699 - 1710.
182. **Shimoda, N., A. Toyoda-Yamamoto, J. Nagamine, M. Katayami, Y. Sakagami, and Y. Machida** 1990. Control of expression of *Agrobacterium vir* genes by synergistic actions of phenolic signal molecules and monosaccharides. *Proc. Natl. Acad. Sci. USA*. **87**:6684-6688.
183. **Shurvinton, C. E., L. Hodges, and W. Ream** 1992. A Nuclear Localization Signal and the C-Terminal Omega Sequence in the *Agrobacterium tumefaciens* VirD2 endonuclease are important for Tumor Formation. *Proc. Natl. Acad. Sci. USA*. **89**:11837 - 11841.
184. **Silversmith, R. E., and R. B. Bourret** 1999. Throwing the switch in bacterial chemotaxis *Trends in Microbiology*. **7**.
185. **Simon, R., J. Quandt, and W. Klipp** 1989. New derivatives of transposon Tn5 suitable for mobilization of replicons, generation of operon fusions and induction of genes in Gram-negative bacteria. *Gene*. **80**:161-169.
186. **Smit, G., T. J. J. Logman, M. E. T. I. Boerrigter, J. W. Kijne, and B. J. J. Lugtenberg** 1989. Purification and partial characterization of the *Rhizobium leguminosarum* biovar *viciae* Ca<sup>2+</sup>-dependent adhesin, which mediates the first

- step in attachment of cells of the family *Rhizobiaceae* to plant root hair tips J. Bacteriol. **171**:4054-5062.
187. **Song, Y., M. Shibuya, Y. Ebizuka, and U. Sankara** 1991. Synergistic action of phenolic signal compounds and carbohydrates in the induction of virulence gene expression of *Agrobacterium tumefaciens*. Chem. Pharm. Bull. **39**:2613-2616.
  188. **Song, Y., M. Shibuya, Y. Ebizuka, and U. Sankawa** 1991. Identification of plant factors inducing virulence gene expression in *Agrobacterium tumefaciens*. Chem. Pharm. Bull. **39**:2347-2350.
  189. **Sourjik, V., and R. Schmitt** 1996. Different Roles of CheY1 and CheY2 in the Chemotaxis of *Rhizobium meliloti*. Molecular Microbiology. **22**:427-436.
  190. **Sourjik, V., W. Sterr, J. Platzer, I. Bos, M. Haslbeck, and R. Schmitt** 1998. Mapping of 41 Chemotaxis, Flagellar and Motility Genes to a Single Region of the *Sinorhizobium meliloti* Chromosome. Gene. **223**:283-290.
  191. **Stachel, S. E., E. Messens, M. Van Montagu, and P. Zambryski** 1985. Identification of the signal molecules produced that activate T-DNA transfer in *Agrobacterium tumefaciens* Nature. **318**:624-629.
  192. **Stachel, S. E., E. W. Nester, and P. C. Zambryski** 1986. A plant cell factor induces *Agrobacterium tumefaciens vir* gene expression. Proc. Natl. Acad. Sci. USA. **83**:379-383.
  193. **Stachel, S. E., B. Timmerman, and P. C. Zambryski** 1986. Generation of single-stranded T-DNA molecules during the initial stages of T-DNA transfer from *Agrobacterium tumefaciens*. Nature. **322**:706-712.
  194. **Stewart, R. C.** 1993. Activating and inhibitory mutations in the regulatory domain of CheB, the methylesterase in bacterial chemotaxis. J. Biol. Chem. **268**:1921-1930.
  195. **Stock, A. M., E. Martinez-Hackert, B. F. Rasmussen, A. H. West, J. B. Stock, D. Ringe, and G. A. Petsko** 1993. Structure of the Mg<sup>2+</sup>-bound form of CheY and mechanism of phosphoryl transfer in bacterial chemotaxis. Biochem. **32**:13375-13380.
  196. **Stock, A. M., and S. L. Mowbray** 1995. Bacterial Chemotaxis - A Field in Motion. Current Opinion in Structural Biology. **5**:744 - 751.

197. **Stock, J.** 1990. Role of protein carboxyl methylation in bacterial chemotaxis., p. 275-284. *In* W. K. Paik, and S. Kim (eds), Protein Methylation. CRC Press Inc., Boca Raton Fl.
198. **Stock, J. B., G. S. Lukat, and A. M. Stock** 1991. Bacterial chemotaxis and the molecular logic of intracellular signal transduction networks. *Annu. Rev. Biophys. Biophys. Chem.* **20**:109-136.
199. **Stock, J. B., A. M. Stock, and J. M. Mottonen** 1990. Signal transduction in bacteria. *Nature.* **344**:395-400.
200. **Stock, J. B., M. G. Surette, W. R. McCleary, and A. M. Stock** 1992. Signal transduction in bacterial chemotaxis. *J. Biol. Chem.* **267**:19753-19756.
201. **Stock, J. B., and M. Surette** 1996. Chemotaxis, in *Escherichia coli* and *Salmonella typhimurium*, 2nd ed, vol. 1. ASM press, washington, D.C.
202. **Sundberg, C., L. Meek, K. Carroll, A. Das, and W. Ream** 1996. VirE1 Protein Mediates Export of the Single-Stranded DNA-Binding Protein VirE2 From *Agrobacterium tumefaciens* into Plant Cells. *Journal of Bacteriology.* **178**:1207-1212.
203. **Surette, M. G., M. Levit, Y. Liu, G. Lukat, E. G. Ninfa, A. Ninfa, and J. B. Stock** 1996. Dimerization is required for the activity of the protein histidine kinase CheA that mediates signal transduction in bacterial chemotaxis. *Journal of Biological Chemistry.* **271**:939-945.
204. **Suzuki, T., and T. Iino** 1981. Role of the *flaR* gene in hook formation in *Salmonella* spp. *J. Bacteriol.* **148**:973-979.
205. **Taylor, B. L., and M. S. Johnson** 1993. Universal themes of signal transduction in bacteria, p. 3-15. *In* J. Kurjan, and B. L. Taylor (eds), Signal Transduction: Prokaryotic and Simple Eukaryotic Systems. Academic Press, Inc., San Diego.
206. **Taylor, B. L., and J. W. Lengeler** 1990. Transductive coupling by methylated transducing proteins and permeases of the phosphotransferase system in bacterial chemotaxis, p. 69-90. *In* R. C. Aloia, C. C. Curtain, and L. M. Gordon (eds), Membrane Transport and Information Storage. Alan R. Liss Inc., New York.
207. **Taylor, B. L., and I. B. Zhulin** 1998. In Search of Higher Energy: Metabolism-Dependent Behaviour in Bacteria. *Molecular Microbiology.* **28**:683 - 690.

208. **Thanassi, D. G., and S. J. Hultgren** 2000. Multiple pathways allow protein secretion across the bacterial outer membrane *Current Opinion in Biology*. **12**:420-430.
209. **Tinland, B.** 1996. The Integration of T-DNA into Plant Genomes. *Trends in Plant Science*. **1**:178 - 184.
210. **Tinland, B., B. Hohn, and H. Puchta** 1994. *Agrobacterium tumefaciens* Transfers Single-Stranded Transferred DNA (T-DNA) into the Plant Cell Nucleus. *Proc. Natl. Acad. Sci. USA*. **91**:8000-8004.
211. **Tinland, B., Z. Koukolikova-Nicola, M. N. Hall, and B. Hohn** 1992. The T-DNA-linked VirD2 protein contains two distinct functional nuclear localization signals. *Proc. Natl. Acad. Sci. USA*. **89**:7442 - 7446.
212. **Tisa, L. S., and J. Adler** 1992. Calcium ions are involved in *Escherichia coli* chemotaxis. *Proc. Natl. Acad. Sci. USA*. **89**:11804-11808.
213. **Toker, A. S., M. Kihara, and R. M. Macnab** 1996. Deletion Analysis of the FlhM Flagellar Switch Protein of *Salmonella typhimurium*. *Journal of Bacteriology*. **178**:7069-7079.
214. **Turk, S. C. H. J., L. S. Melchers, H. den Dulk-Ras, A. J. G. Regensburg-Tuink, and P. J. J. Hooykaas** 1991. Environmental conditions differentially affect *vir* gene induction in different *Agrobacterium* strains. Role of the VirA sensor protein. *Plant Mol. Bio*. **16**:1051-1059.
215. **Turk, S. C. H. J., R. P. Van Lange, E. Sonneveld, and P. J. J. Hooykaas** 1993. The Chimeric VirA-Tar Receptor Protein is Locked into a Highly Responsive State. *Journal of Bacteriology*. **175**:5706-5709.
216. **Valdivia, R. H., L. Wang, and S. C. Winans** 1991. Characterization of a Putative Periplasmic Transport System for Octopine Accumulation Encoded by *Agrobacterium tumefaciens* Ti Plasmid pTiA6. *Journal of Bacteriology*. **173**:6398-6405.
217. **Van Larebeke, N., G. Engler, M. Holsters, S. Van den Elsacker, I. Zaenen, R. A. Schilperoort, and J. Schell** 1974. Large plasmid in *Agrobacterium tumefaciens* essential for crown gall-inducing ability *Nature*. **252**:169-170.
218. **Vogel, A. M., and A. Des** 1992. Mutational analysis of *Agrobacterium tumefaciens virD2*; Tyrosin 29 is essential for endonuclease activity *J. Bacteriol*. **174**:303-308.

219. **Vogler, A. P., M. Homma, V. M. Irikura, and R. M. Macnab** 1991. *Salmonella typhimurium* mutants defective in flagellar filament regrowth and sequence similarity of FliI to F<sub>0</sub>F<sub>1</sub>, vacuolar, and archaebacterial ATPase subunits. *J. Bacteriol.* **173**:3564-3572.
220. **Vogler, A. P., and J. W. Lengeler** 1987. Indirect role of adenylate cyclase and cyclic AMP in chemotaxis to phosphotransferase system carbohydrates in *Escherichia coli* K-12. *J. Bacteriol.* **169**:593-599.
221. **Welch, M., N. Chinardet, L. Mourey, C. Birck, and J. Samama** 1998. Structure of the CheY-Binding Domain of Histidine Kinase CheA in Complex with CheY. *Nature Structural Biology.* **5**:25 - 29.
222. **Welch, M., K. Oosawa, S.-I. Aizawa, and M. Eisenbach** 1993. Phosphorylation-dependent binding of a signal molecule to the flagellar switch of bacteria. *Proc. Natl. Acad. Sci. USA.* **90**:8787-8791.
223. **Winans, S. C.** 1991. An *Agrobacterium* two-component regulatory system for the detection of chemicals released from plant wounds. *Mol. Microbiol.* **5**:2345-2350.
224. **Winans, S. C.** 1992. Two-way chemical signaling in *Agrobacterium*-plant interactions. *Microbiol. Rev.* **56**:12-31.
225. **Winans, S. C., R. A. Kerstetter, and E. W. Nester** 1988. Transcriptional regulation of the *virA* and *virG* genes of *Agrobacterium tumefaciens* *J. Bacteriol.* **170**:4047-4054.
226. **Winans, S. C., N. J. Mantis, C. Y. Chen, C. H. Chang, and D. C. Han** 1994. Host recognition by the VirA, VirG two-component regulatory proteins of *Agrobacterium tumefaciens* *Research Microbiology.* **145**:461-473.
227. **Woodcock, D. M., P. J. Crowther, J. Doherty, S. Jefferson, E. DeCruz, M. Noyer-Weidner, S. S. Smith, M. Z. Michael, and M. W. Graham** 1989. Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. *Nucleic Acids Research.* **17**:3469-3478.
228. **Wright, E. L.** 1999. PhD. Thesis. University of Durham.
229. **Wu, J., J. Li, G. Li, D. G. Long, and R. M. Weis** 1996. The chemotaxis receptors from *Escherichia coli* bind to the methyltransferase (CheR) at a site distinct from the sites of methylation., p. 313. *In* Abstracts of the General Meeting of the American Society for Microbiology.

230. **Wu, J. G., J. Y. Li, G. Y. Li, D. G. Long, and R. M. Weis** 1996. The receptor binding site for the methyltransferase of bacterial chemotaxis is distinct from the sites of methylation. *Biochemistry*. **35**:4984-4993.
231. **Yadav, N. S., J. Vanderleyden, D. R. Bennet, W. M. Barnes, and M.-D. Chilton** 1982. Short direct repeats flank the T-DNA on a nopaline Ti-plasmid *Proc. Natl. Acad. Sci. USA*. **79**:6332-6326.
232. **Yong, G. M., D. H. SCHMIL, and V. L. MILLER** 1999. A new pathway for secretion of virulence factors by bacteria: the flagellar export apparatus functions as a protein- secretion system *Proc. Natl. Acad. sci. USA*. **96**:6456-6461.
233. **Zambryski, P., J. Tempe, and J. Schell** 1992. Chronicles from the *Agrobacterium*-plant cell DNA transfer story *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **43**.
234. **Zanker, H., G. Lurz, U. Langridge, P. Langridge, D. Kreusch, and J. Schroder** 1994. Octopine and Nopaline Oxidases from Ti Plasmids of *Agrobacterium tumefaciens*: Molecular Analysis, Relationship, and Functional Characterization. *Journal of Bacteriology*. **176**:4511-4517.
235. **Zanker, H., J. Von Lintig, and J. Schroder** 1992. Opine Transport Genes in the Octopine (*occ*) and Nopaline (*noc*) Catabolic Regions in Ti Plasmids of *Agrobacterium tumefaciens*. *Journal of Bacteriology*. **174**:841-849.
236. **Zhu, J., P. M. Oger, S. barbara, H. P. J. J., S. K. Farrand, and S. C. Winans** 2000. The Bases of Crown Gall Tumorigenesis *Journal of Bacteriology*. **182**:3885-3895.
237. **Zhu, Y., L. S. Pierson III, and M. C. Hawes** 1997. Induction of Microbial genes for pathogenesis, and symbiosis by chemicals from root border cells *plant Physiology*. **11**:1691-1698.
238. **Zhuang, W. Y., and L. shapiro** 1995. *Caulobacter* FliQ and FliR membrane proteins, required for flagllar biogenesis and cell division, belong to a family of virilence factor export proteins *Journal of Bacteriology*. **177**:343-356.
239. **Zupan, J., T. R. Muth, O. Draper, and P. Zambryski** 2000. The transfer of DNA from *Agrobacterium tumefaciens* into plants: a feast of fundeamental insights *The plant Journal*. **23**:11-28.



240. **Zupan, J., D. Ward, and P. Zambryski** 1998. Assembly of the VirB transport complex for DNA transfer from *Agrobacterium tumefaciens* to plant cells. **1**:649-655.
241. **Zupan, J. R., V. Citovsky, and P. Zambryski** 1996. *Agrobacterium* VirE2 Protein Mediates Nuclear Uptake of Single-Stranded DNA in Plant Cells. Proc. Natl. Acad. Sci. USA. **93**:2392-2397.
242. **Zupan, J. R., and P. Zambryski** 1995. Update on plant transformation - transfer of T-DNA from *Agrobacterium* to the plant cell. **107**:1041-1047.

